Using metabolics for the exploration of diet and health in twins

Pallister, Tess Kathleen Jean

Awarding institution:
King's College London

USING METABOLOMICS FOR THE EXPLORATION OF DIET AND HEALTH IN TWINS

Tess Kathleen Jean Pallister / 1239451

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Department of Twin Research and Genetic Epidemiology
King's College London

2017

Abstract

Metabolomics is an exciting area of research obtaining the chemical and metabolic signals from biosamples. Food consumption is strongly linked to metabolism making the metabolome an ideal phenotype for identifying biomarkers and helping refine and explore diet-disease associations. However, the metabolome is highly complex and influenced by many factors, such as age, disease, genetics and gut microbiota. Discordant monozygotic twins may provide a strong model for confirming association findings as they are matched for age, sex and the baseline genetic sequence.

In this thesis, I explored the potential and applicability of the metabolome in nutritional research in two main areas: for identifying biomarkers of food exposure and further investigating the relationship of food intake with indicators of health. Firstly, I examined metabolomics profiles associated with self-reported food intakes and dietary patterns and confirmed these associations using the co-twin control method. I then identified top metabolite markers of food group intakes, and created and tested metabolite scores using multiple-methods. I searched for markers of gut microbiome diversity, an emerging indicator of health, by examining the association of alpha-diversity with blood metabolomics profiles and the relationship with diet and the metabolic syndrome. In the final chapter, I created a dietary score predictive of visceral fat mass, a strong risk factor for cardio-metabolic disease, and examined the degree to which the relationship between diet and visceral fat mass is mediated by associated metabolites and microbiome taxa. Throughout each chapter I used discordant monozygotic twins to validate top results. Overall, my findings show that metabolomics is a highly versatile tool for advancing nutrition research from biomarker identification through to untangling the impact of dietary exposures on indicators of metabolic health and the gut microbiome.

Table of Contents

Table of Figures

Table of Tables

Acknowledgements

I'd like to thank first and foremost my supervisors Professor Tim Spector and Dr Cristina Menni. Through their encouragement and guidance I have made my way through this experience in a smooth and timely manner. They also allowed me the freedom to explore topics under my own initiative which I am grateful for.

I'd like to thank Dr Ana Valdes for her guidance on the exploration of diet and visceral fat mass. I believe I learned a lot from that exercise and her expertise was appreciated.

I appreciate all of the other members for the DTR for their helpful advice and data provision throughout my PhD.

Foremost I would like to express gratitude to my parents for their constant support in this endeavour.

Abbreviations

Abbreviation	Meaning
ACSM2A	acyl-CoA synthetase medium-chain family member 2A
ACSM5	acyl-CoA synthetase medium-chain family member 5
AHR	Aryl hydrocarbon receptor
AUC	Area under the receiver operating characteristic curve
BCAA	branched-chain amino acids
BMI	Body mass index
CMPF	3-carboxy-4-methyl-5-propyl-2-furanpropanoate
CPS1	Carbamoyl-phosphate synthase 1
CVD	Cardiovascular disease
CYP1A1	Cytochrome P450 family 1 subfamily A polypeptide 1
CYP1A2	Cytochrome P450 family 1 subfamily A member 2
DHA	Docosahexaenoate
DHEA	Dehydroepiandrosterone
DNA	Deoxyribonucleic acid
DXA	Dual-energy X-ray absorptiometry
DZ	Dizogotic
EFA	Essential fatty acid
EPA	Eicosapentaenoate
EPIC	European Prospective Investigation into Diet and Cancer
F&V	Fruit and vegetable diet pattern
FA	Fatty acid
FADS1	Fatty acid desaturase 1
FBG	Fasting plasma glucose
FD	Fermented dairy
FF	Fried and fast foods
FFQ	Food frequency questionnaire
GAT2	gamma-Aminobutyric acid transporter 2
GWAS	Genome-wide association study
HAO2	Hydroxyacid oxidase 2
HDL	High density lipoprotein cholesterol
KORA	Cooperative health research in the Region of Augsburg
MACS	Medium-chain acyl-CoA synthetase
MDS	Mediterranean diet score
MetS	Metabolic syndrome
MUFA	Monounsaturated fatty acid
MZ	Monozygotic
NAFLD	Non-alcoholic fatty liver disease
NASH	Non-alcoholic steatohepatitis
NAT2	N-acetyltransferase 2
NHANES	National Health and Nutrition Examination Survey
NS	Not significant
NSP	Non-starch polysaccharide
Ost-alpha	Organic solute transporter alpha
Ost-beta	Organic solute transporter beta
OTU	Operational taxonomic unit
PCA	Principal components analysis

PPARα	Peroxisome proliferator-activated receptor alpha
PPARγ	Peroxisome proliferator-activated receptor gamma
PUFA	Polyunsaturated fatty acid
ROC	Receiver operating characteristic curve
SD	Standard deviation
SE	Standard error
SLC16A9	Solute carrier family 16 member 9
SLC51A	Solute carrier family 51 alpha subunit
SLC5A11	Solute carrier family 5 member 11
SLC6A13	Solute carrier family 6 member 13
SLC6A20	Solute carrier family 6 member 20
SLE	Systemic lupus erythematosus
SNP	Single nucleotide polymorphism
SSB	Sugar-sweetened beverage
SULT2A1	Sulfotransferase family 2A member 1
TG	Triglycerides
TMAO	Trimethylamine N-oxide
VF	Visceral fat
VFM	Visceral fat mass
WC	Waist circumference
WDR85	diphthamide biosynthesis 7
WG	Whole grain

Chapter 1 Introduction

In this chapter I discuss twin studies that have examined the heritability of reported food and nutrient intakes and dietary patterns. I then introduce the usage of monozygotic (MZ) twins for recent nutriomics studies including specifically the metabolome and gut microbiome. Finally, I briefly introduce the usage of twins combined with nutriomic methods for the exploration of the effect of diet on metabolic disease.

Part of this work has been published as a literature review in *Nutrition Research Reviews* (Pallister et al., 2014).

Early nutrition studies on twins and multi-foetal pregnancies focused on the side effects of the pregnancy, achieving adequate maternal nutrition during gestation, and on the implications for the health of the growing foetuses (Brown and Carlson, 2000). The womb, however, is a solitary internal environment, where nutritional parameters can be measured with more ease. Once twins enter the world, where myriad of external environmental influences on different phenotypes can be assessed, twins provide us with the unique ability to accurately determine heritability of certain complex traits, such as dietary phenotypes. In this context, dietary phenotypes broadly refer to the observable and measurable foods, and their constituents, consumed by individuals or groups. Indeed, twin studies have proven that genetic makeup plays a significant role in dietary intakes (Rankinen and Bouchard, 2006). Moreover, twins have the advantage of limiting inter-individual variability as they are raised in a similar environment and have matched age, genes and sex (for monozygotic [MZ] pairs). They therefore represent a natural matched case-control experiment.

Omics technologies allow for a systems approach to nutrition research, that encompasses the primary level of DNA sequence (genomics), gene expression (transcriptomics) and epigenome, to the intermediate phenotypes (proteome, metabolome, and microbiome), finally to clinical endpoints (Kussmann et al., 2006). Food constituents may influence each of these areas, but their effect is difficult to ascertain due to imprecise dietary assessment methods. Omics techniques may help ameliorate this issue through food intake biomarker discovery within the intermediate phenotypes, primarily the metabolome (Zivkovic and German, 2009). Once biomarkers of intake

have been validated, in conjunction with currently utilized intake and known biomarker measures, reliable associations between genomic and transcriptomic data can be made. This can allow for identification of groups at risk for poor dietary patterns, gene-diet interactions, and diet-disease associations, which combined could conjure an accurate depiction of a “healthy” phenotype (Kussmann et al., 2006), although this is limited by currently used methods of dietary assessment.

Presently, epidemiological studies rely mainly on subjective reporting to obtain dietary information. These include prospective methods such as weighted and estimated food records, and retrospective methods, including 24-hour recalls and food frequency questionnaires (FFQs). Each of these methods carry their own strengths and weaknesses, which have been reviewed by Shim et al. (2014). Weighted records and estimated food records carry a higher participant burden, and use more time and costs, though they provide more accurate data as they do not rely on participant memory (i.e. recall bias). 24 hour recalls commonly require an interviewer (though computer-based methods are becoming more common-place), they are limited by possible recall bias and interviewer bias and are expensive and time-consuming. Both food records and 24 hour recalls require multiple days of evaluation to assess usual food intake. FFQs are the most commonly used dietary assessment method for population-based studies as they have a lower participant burden and resources and indicate habitual food consumption, though are limited to a set list of foods and have low accuracy due to recall bias. With these limitations in mind, identifying novel dietary biomarkers using multi-omics will be challenging and require confirmation and evaluation by dietary intervention studies.

In this introduction, I first discuss the usage of twins for generating heritability estimates of food intake phenotypes, including energy and macronutrient intakes, dietary patterns and specific food group intakes. I then highlight the value of discordant MZ pairs, for furthering nutrition research through discovery and validation of biomarkers of intake and health status in conjunction with cutting-edge omics technologies, with metabolomics being the focal point.

1.1 Heritability of food intake phenotypes: The usage of twins for heritability estimates of intakes

The classic twin design allows for a natural experiment that exploits the difference in genetic relatedness between MZ and dizygotic (DZ) twins to estimate the degree to which phenotypic variability, such as food or nutrient intake, is explained by genetic and environmental factors. Heritability is defined as the degree of total phenotypic variance due to genetic variation and is relevant only to groups or populations and not at the individual level (Visscher et al., 2008).

1.1.1 Heritability of energy and macronutrient intakes

Accurate assessments of the genetic influence on energy intakes is crucial as these provide a baseline assessment for which genes influence total food intakes (de Castro, 1993a) and therefore justify genotype-specific dietary intervention strategies. Furthermore, the degree to which macronutrient intakes are genetically determined may have significant implications for the role of elevated intakes of energy dense foods (i.e. those containing high concentrations of fat and refined carbohydrates) to the obesity epidemic (Guyenet and Schwartz, 2012). Therefore, multiple twin studies have evaluated these components through various dietary assessment methods (**Figure 1-1a**; **Appendix A Table 1**).

Studies of different size and quality have estimated heritability of different energy and macronutrient intakes to lie between 8% and 70% (Liu et al., 2013, Pimpin et al., 2013, de Castro, 1993a, Heller et al., 1988, Wade et al., 1981, Hasselbalch et al., 2008, Aden et al., 1979, Hur et al., 1998). Beginning from a young age, infant twins have showed that the genetic component provides a small, albeit significant, effect on energy and macronutrient intakes (8% to 12%) (Pimpin et al., 2013). As children age and they become less dependent on their parents, the genetic effect on energy and macronutrient intakes appears to intensify, as suggested by findings in 11-13 year-old US twins (31% to 48%) (Liu et al., 2013). This pattern is similar to other health-related phenotypes, including BMI (Dubois et al., 2012), which has been recently confirmed through a more complex longitudinal genome-wide complex trait analysis for determining DNA-based heritability (Llewellyn et al., 2014). Heritability of dietary energy and macronutrients in adults has been found to vary widely: for energy (32% to 65%), for fats (35% to 53%), carbohydrates (25% to 67%), and proteins (28% to

70%) (Hasselbalch et al., 2008, Hur et al., 1998, de Castro, 1993a, Wade et al., 1981). An extensive study of adult Danish twins showed inheritance of energy and energy-adjusted macronutrient intakes to range from 28% to 55% (Hasselbalch et al., 2008). Although, genetic effects on macronutrient intakes were more substantial in men (49% to 55%) than women (28% to 36%), they were not different between macronutrients. In support of this, two other twin studies suggested the same genetic mechanisms governing energy intakes (Guyenet and Schwartz, 2012) influence macronutrient intakes (de Castro, 1993b, Faith et al., 1999).

The lack of evidence to date for independent sets of genes influencing macronutrient intakes from those genes governing energy intakes has focussed attention on whether dietary energy density (kilocalories per gram of food) is genetically determined, with one study suggesting substantial genetic influence (de Castro, 2006), and another not (Hasselbalch et al., 2008). Despite this, in the latter Danish study, factors related to dietary energy density, including fibre, glycaemic index and the glycaemic load, were found to be significantly heritable both in women (49%, 36% and 33%, respectively) and men (41%, 30%, and 25%, respectively). The dependence of these phenotypes on the constituents of the whole diet suggests genetic effects on these phenotypes may be reflective of variable dietary patterns.

1.1.2 Heritability of dietary patterns

Dietary patterns are identified by two primary methods: empirically-derived or *a priori* (Hu, 2002). Empirically-derived dietary patterns employ statistical methods to identify natural groupings of intakes of food items (e.g., principal component analysis or cluster analysis (Hu, 2002)). *A priori* dietary patterns are generated based on adherence to diet parameters previously associated with health outcomes or physiological states, such as the Mediterranean diet score. A key advantage to using dietary patterns is the translatable results that encompass the degree to which genetics influence the global diet.

Two twin studies on children using empirically-derived dietary patterns generated highly inconsistent heritability estimates despite reasonable sample size (**Appendix A Table 2**) (Breen et al., 2006, Faith et al., 2008). In a study of US 7 year olds (*n*=792) heritability estimates of food pattern intakes obtained by 24-hour recalls were stronger in boys than girls, ranging from 12% to 79% and 20% to 56%, respectively (Faith et al., 2008). In another study of 4 to 5 (SD: 0.3) year-old

UK twin-pairs (*n*=214) heritability of food preference was measured through the use of a modified 95-item "liking" food frequency questionnaire (FFQ) administered to parents (Breen et al., 2006). Heritability ranged from 20% (desserts) to 78% (meat and fish) on food liking-disliking groups. It is difficult to make valid comparisons between these two studies as the dietary assessment methods used measure different outcomes (i.e. intakes versus liking), although it has been found that exposures to foods generally encourages liking and intakes in children (Wardle et al., 2001). This suggests environmental influences may override genetic predispositions to food pattern intakes in children, which has implications for combating food neophobia, a highly heritable and perhaps evolutionarily important trait (Cooke et al., 2007).

Contrary to studies in children, studies in adults have generated relatively stable heritability estimates for food pattern intakes across patterns and genders. "Healthy" dietary patterns, characterised by high intake of vegetables, fruits and whole grains and low intake of fatty foods of animal origin and simple carbohydrates, have heritability estimates ranging from 33% to 54%, while "unhealthy" patterns ranged from 33% to 50% (**Appendix A Table 2**) (van den Berg et al., 2013, Keskitalo et al., 2008, Teucher et al., 2007, Gunderson et al., 2006, van den Bree et al., 1999). Furthermore, a study which utilized a twins of mistaken zygosity approach (a useful method to control for potential equal environmental assumption bias (Scarr, 1968)) in US female twins (*n*=700), found additive genetic effects to account for 50% of the variability in healthy pattern intakes, while an unhealthy diet pattern was not significantly heritable (Gunderson et al., 2006). A healthy diet is more phenotypically similar to diets humans have consumed through millennia and evolved with, it is only very recently that high energy dense and processed foods have been readily available, as such it makes sense the variation in this diet may be more genetically influenced in some groups (Breslin, 2013). One study conducted on an older age group (≥ 50 years) determined additive genetic effects to account for 33% of the variability of consumption frequency for both healthy and unhealthy patterns (van den Bree et al., 1999). A pattern of increased dietary variety in modern societies may contribute to the obesity epidemic through ready availability of highly processed, calorific foods compared to healthier, natural choices such as fruits and vegetables (McCrory et al., 1999). In middle age to elderly twins from the "Virginia 30,000" twin study (N = 5,543) dietary variety seeking was up to 30% heritable, supporting the notion that our environment is an important driver of dietary variety (Scheibehenne et al., 2014).

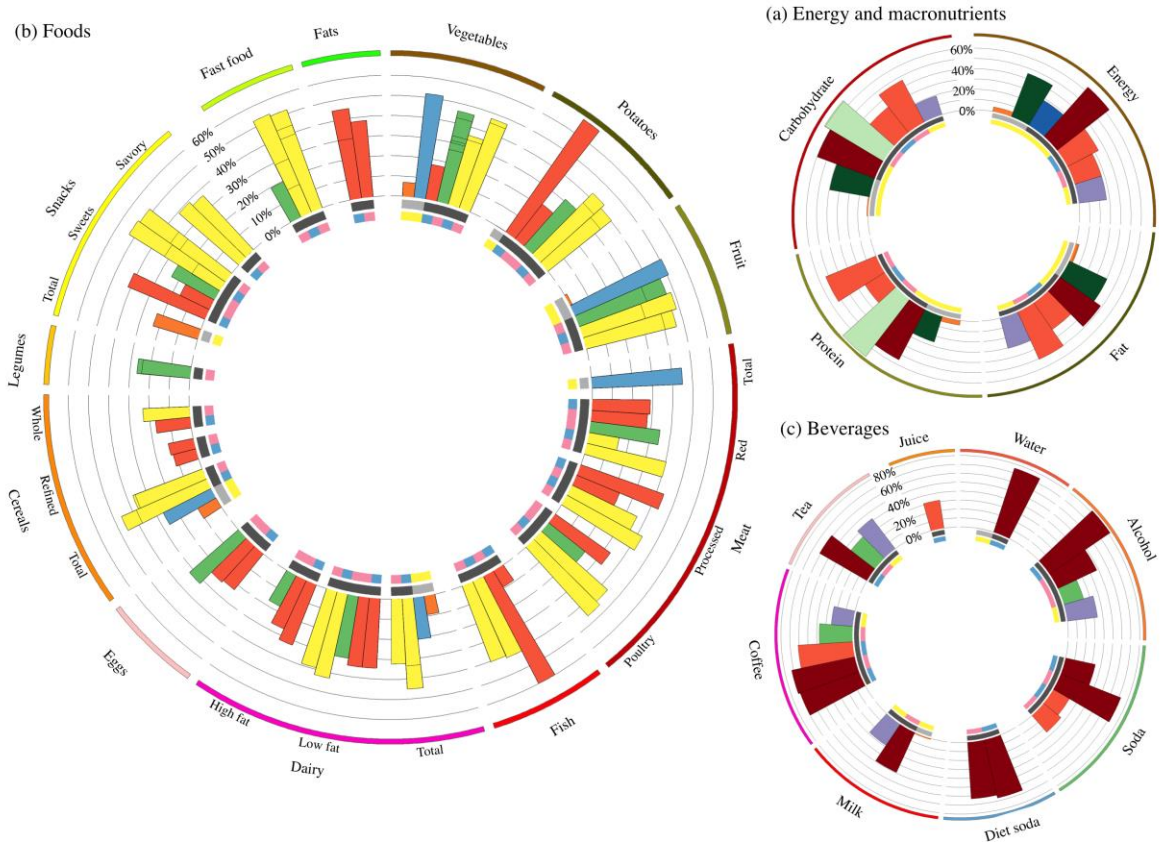


Figure 1-1. Estimated heritability of nutrients (a), food (b) and beverage (c) intakes from twin studies
Intake heritabilities presented are significant findings from previous twin studies. Heritability histograms are color coded according to study. From clockwise, histograms are grouped according to age; the first line below the histogram denotes this: light grey, children; dark grey, adults. Within each age group, histograms were grouped according to accuracy of the dietary assessment method used, from most accurate (e.g. 2 day buffet style meal intervention) to least accurate (e.g. 67-item FFQ). The second line below the histograms indicates gender: pink, female; blue, male; yellow, combined. *Modified from Pallister et al.(Pallister et al., 2014)*

1.1.3 Individual foods and food group heritabilities

A number of recent studies have estimated the genetic effect of food and beverage intakes in humans (**Figures 1-1b** and **1-1c; Appendix A Tables 3** and **4**) generating wide-ranging heritability estimates across food groups, particularly for children (Pimpin et al., 2013, Fildes et al., 2014). Findings from the UK Gemini cohort of infants and children aged 21 months to 3 years have shown age effects on heritability of food group intakes are evident even in the early years (Fildes et al., 2014, Pimpin et al., 2013). Intakes for 21 month-old children were minimally heritable for food groups ranging from 9% (potatoes) to 17% (dairy). Whereas preference for food groups were highly heritable in 3 year-old children, ranging from 27% (dairy) to 54% (vegetables). It should be noted that heritability estimates in the older children were calculated from age- and sex-adjusted residuals, an approach which may have prevented inflation of shared environmental effects (McGue and Bouchard, 1984). The study on infants did not use this approach, perhaps resulting in an underestimation of the genetic effects on food group intakes.

Recent studies in adults have also confirmed there is a genetic component to many food type intakes. Danish twins showed a large degree of variation in calculated heritabilities overall ranging from 17% (fish) to 68% (potato) in men, and 20% (whole grain cereal) to 61% (fish) in women (Hasselbalch et al., 2008). Similar results were seen in UK female twins where heritabilities ranged from 8% (refined grains) to 46% (garlic) (Teucher et al., 2007). Furthermore, the grouping of fruit and vegetable sources generated the highest heritability estimate in this cohort (49%). These authors proposed taste perception may be a key driver of the predicted genetic influence as foods characterised by strong tastes are the cornerstone of these groups. This observation extends to consistently substantial heritabilities for coffee as well as alcohol intakes, ranging from 29% to 73% and 28% to 82%, respectively (Hasselbalch et al., 2008, Teucher et al., 2007, Hur et al., 1998, de Castro, 1993b). Although, the pharmacological properties of both must be considered as alcohol dependence, as well as caffeine consumption-related traits, are both highly heritable (30% to 70%, and 36% to 58%, respectively) (Agrawal and Lynskey, 2008, Yang et al., 2010a) and may drive intakes up in a way independent of taste preference or a biological requirement for these items.

Taste is the strongest determinant of food choices (Feeney et al., 2011), and a recent twin study showed multiple chemosensory facets of numerous food compounds to be highly heritable

and associated with particular gene variants (Knaapila et al., 2012). Future heritability studies of food intakes may shift the focus to food groupings characterized by specific tastes.

1.1.4 Nutritional phenotype heritability conclusion

In summary, twin studies have shown food and macronutrient intakes to be genetically determined. However, heritability estimates are often inconsistent between studies for multiple reasons. First, heritability is only specific to the population studied, which can vary greatly, due to the vast country specific food environments. Indeed, if a food is not readily available in a population there is no capacity for intakes to be genetically determined. Second, important evidence is available for familial and community influence on food intake (such as parenting styles, food exposures, socioeconomic status) which have an important place for battling current issues, such as childhood obesity (Berge, 2009) and food neophobia (Anzman-Frasca et al., 2012), and diabetes (Pollard et al., 2014). Finally, methods for determining heritability are not standardised, for instance the inclusion/exclusion of variables from the model (e.g. age, country), the segregation/aggregation of sexes, and the different methods for obtaining and analysing dietary data tend to be unique for each study. The same genetic mechanisms governing energy intakes are suspected to influence macronutrient intakes. However, recent studies have assessed the impact of gene variants encoding proteins related to the complex energy intake regulatory systems under central nervous system control (Guyenet and Schwartz, 2012) on preference for macronutrient types. These studies are few on twins (Hasselbalch et al., 2010, Bouchard-Mercier et al., 2012). Heritability estimates are more limited for macronutrient subtype intakes (e.g. the fatty acid profile (Hur et al., 1998, Heller et al., 1988)), different age groups or changing physiological states (e.g. pregnancy), and socially driven diets (e.g. veganism). Macronutrient profiles do not account entirely for food pattern preferences that may be more determined by genetic makeup. Variation in diet patterns of adults are to a relatively high degree attributable to genetics (van den Berg et al., 2013, Keskitalo et al., 2008, Teucher et al., 2007, Gunderson et al., 2006, van den Bree et al., 1999). Furthermore, specific food type intakes that characterise a healthy diet pattern tend to be highly heritable in adults and in turn consist of foods that elicit distinctive taste responses (e.g. garlic, fruit and vegetables) (Teucher et al., 2007).

The substantial heritability of multiple dietary phenotypes supports the concept of personalised dietary recommendations, the generation of which has so far primarily relied on the exploration of an individual's genetic makeup. However, on a molecular level, the search for genotypes predisposing individuals to particular dietary intakes is very much in its infancy. Nonetheless, the evidence generated by twin heritability studies can lead scientists to dietary phenotypes that are consistently highly heritable and demand further exploration through GWAS and candidate gene studies, following which twins can aid the determination of the heritable contribution of specific singly nucleotide polymorphisms (SNPs) to the dietary phenotype (as used previously for human height (Yang et al., 2010b)).

A key issue which has long plagued nutrigenetic research arises from the inaccuracies of self-reported diet intake and food preference data and as a result, the hunt for dietary biomarkers, objective measures of dietary intakes and nutritional status, has been hastened. These may allow for more accurate heritability calculations for food and nutrient intakes as has recently been shown for salt intakes using urinary sodium excretion, an established, valid biomarker of salt intake (Kho et al., 2013), and could potentially be used in genetic association studies. However, the usage of biomarkers is not confined to genetic studies. It is becoming increasingly clear that for the establishment and refinement of personalised nutrition, in combination with genetic information, a whole organism, multi-systems approach must be undertaken, where the usage of biomarkers of food intakes and nutritional status will be essential. Nutrition research is now embarking on a nutriomics era, where emerging high-throughput omics technologies will substantially aid biomarker discovery and "healthy phenotype" definition. As will be explored next, twin studies are progressively taking steps to make a contribution in this way.

1.2 The importance of twins for emerging nutriomic research

This next section will explore emerging dietary studies on twins used in collaboration with high-throughput omics to capture metabolism (metabolomics) and the gut microbiota (microbiome). Integrating these technologies to determine a "healthy" phenotype is the long-term goal of personalised nutrition (van Ommen and Stierum, 2002). A reoccurring issue with this type of analysis is the high degree of inter-individual variation, making associations often inconsistent (Kussmann et al., 2006). As such, the usage of identical twins discordant for dietary factors or

objective measures of nutrition status, matched entirely for age, sex, genetics and partially for early environmental influences, will provide an enhanced method of assessing diet and biological relationships.

Obesity, the greatest and fastest growing health concern in the world at this time, is a highly genetic disorder, influenced by myriad complex factors (Despres et al., 1992). However, rare obese discordant identical twin pairs, provide a unique opportunity to disentangle lifestyle and environmental factors on the human system derangements induced by energy imbalance independently of genetics.

1.2.1 Nutrimetabolomics

Nutritional metabolomics involves the extensive chemical profiling of various tissues completed in a global manner through targeted and non-targeted methods used as a complement to diet- and health-related complex systems approaches. The metabolome gives unique information into the metabolic status of an individual by providing a snapshot of the metabolic processes undertaken in a bodily system, specific organ, tissue or cell, which cannot be identified through measuring gene expression or the proteome. Targeted methods measure a panel or group of chemically defined metabolites. Whereas non-targeted methods aim to detect all measureable chemicals in a sample and therefore have more coverage. Moreover, non-targeted metabolomics has the potential to detect previously unknown metabolites in a sample, increasing the possibility for identifying novel findings. Although, it may be difficult to draw conclusions from such findings. Targeted metabolomics has increased sensitivity over non-targeted methods measuring absolute concentrations as opposed to relative methods used for non-targeted methods. Moreover, targeted methods typically measure metabolites which are biologically well-defined (Patti et al., 2012).

The two primary technologies used to analyse sample metabolomes are mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy. NMR spectroscopy is a quantitative technique that determines information on the solution-state molecular structures in a sample through atom-based nuclear interactions (Marion, 2013). MS-based methods typically separate metabolites first in a sample using chromatography (examples include gas or liquid-based chromatography, and capillary electrophoresis), then metabolites are identified with a mass spectrometer (Feng et al., 2008). Although, some MS methods do not require the initial sample

separation step and metabolites can be quantified directly. Gas or liquid chromatography separate metabolites by time, providing detailed chemical information, though this may require complex chemical preparation that can destroy metabolites (Sellick et al., 2010). Moreover, often techniques (such as GC–MS and LC-MS) are combined as different techniques detect different types of metabolites.

Veenstra (2012) previously summarized the advantages and disadvantages of both methods. The key advantage of MS metabolomics is the sensitivity or the ability to measure very small analytes (within the femtomolar to attomolar range) with accuracy. Although the main disadvantage of MS is less precise quantification due to the MS signal intensity being influenced by the method of sample preparation (Veenstra, 2012), with complex sample preparation also being a disadvantage. Conversely, NMR spectroscopy is highly precise due to the peak area of a compound being proportional to the sample concentration of particular identified nuclei (such as ^{1}H, ^{13}C) (Veenstra, 2012). Moreover, NMR techniques require minimal sample preparation, are non-destructive, and provide rapid results with high reproducibility at a low cost (Markley et al., 2016). The primary disadvantage of NMR-based methods is their lack of sensitivity (Veenstra, 2012), it is also non-discriminatory (Markley et al., 2016).

Nutrition and the metabolome are intimately linked in that nutrients and non-nutrient food constituents supply metabolites, however these contributions make the dietary influence on the metabolome all the more difficult to ascertain (Gibney et al., 2005). Despite this, dietary pattern and food intakes generated through self-reported intake data have identified novel biomarkers within the metabolome (O'Sullivan et al., 2011, Altmaier et al., 2011, Guertin et al., 2014b, Zheng et al., 2014), validating this approach for use in epidemiological studies. However, the genetic influence on metabolite levels is wide-ranging (Suhre et al., 2011), therefore novel usage of the twin model provides an ideal method for determining the dietary impact on metabolites, through segregating the non-genetic component. One nutri-metabolomic study conducted on twins further confirmed self-reported intake associations with metabolites (Menni et al., 2013c). Female UK twins completed 131-item FFQs that generated dietary constituents and intake patterns found to associate with 42 metabolite levels in the larger twin population. Monozygotic twins discordant for dietary intakes were then identified and utilized to replicate results, confirming the utility of this method (Zivkovic and German, 2009).

Lipidomics, a branch of metabolomics, has been used to characterize environmental and lifestyle-induced changes to the global serum lipid profile in 14 healthy, young MZ obesity-discordant twin pairs (Pietilainen et al., 2007). Levels of lysophosphatidylcholines, lipids associated with inflammation (Yang et al., 2005) and atherogenesis (Glass and Witztum, 2001) were found to be elevated in obese co-twins, with concommitant decreases in antioxidant (Wallner and Schmitz, 2011) ether phospholipids. Moreover, despite the young ages of the obese co-twins (24 to 27 years), these lipid profile changes were associated with insulin resistance.

The metabolome is highly complex and inter-individual variation is high for a multitude of reasons (Zulyniak and Mutch, 2011), including age (Menni et al., 2013b), sex (Krumsiek et al., 2015), genetics (Shin et al., 2014), ethnicity (Wikoff et al., 2013) and the gut microbiome (Wikoff et al., 2009), thus factoring solely nutritional intake into nutrimetabolomic studies is problematic and the relative importance of blood and urine levels is unclear. For example, previous analysis on the non-targeted metabolomics TwinsUK dataset showed 22 metabolites to account for 59% of the variance in age (Menni et al., 2013b). In another study conducted on the TwinsUK dataset the heritability of blood metabolite levels ranged from 10-81%, while only 30% of the blood metabolites were not significantly influenced by genetics (Shin et al., 2014). Finally, sex has shown to influence one-third of metabolites, while a network analysis showed certain metabolic pathways to be sex-specific (Krumsiek et al., 2015).

An area of increasing interest for its potential influence on metabolic processes is the gut microbiome. In rodents, the gut microbiota were found to account for 10% of the variability in the plasma metabolome (Wikoff et al., 2009). Although estimates for humans are unknown at this time, regional variation in human metabolomic profiles has been attributed in part to gut microbes (Nicholson et al., 2012). In turn, the gut microbiota carry out processes in the digestion and handling of nutrients, some of which are essential for optimal nutrition, prior to entering the host metabolome (Nicholson et al., 2012). Emerging methods for direct measurement of the intestinal metabolome will help to unravel the microbiome effects on host metabolome (Ursell et al., 2014).

1.2.2 Microbiome

The human gut microbiome is estimated to contain hundreds of bacterial species, with upwards of 536,000 bacterial genes among them (Qin et al., 2010). For years, dietary constituents, such as

pre- and pro-biotics, have been consumed for their proposed health benefits postulated to be through modulation of the gut microbial population (Collins and Gibson, 1999). Population-based studies suggest dietary patterns appear to strongly drive gut microbiome composition. Differences have been observed between Americans and Malawians and Amerindians (Yatsunenko et al., 2012), and consumers of meat- versus plant-based diets (Matijasic et al., 2013, Wu et al., 2011a), findings that are being confirmed by dietary intervention studies (David et al., 2014, O'Keefe et al., 2015). Although the potential for the gut microbiota to drive eating behaviours has also been proposed through modifying cravings and inducing dysphoria (Alcock et al., 2014), for instance individuals showing a desire for chocolate have a distinctive urinary microbial metabolite profile than subjects who were chocolate indifferent, although they consumed identical diets (Rezzi et al., 2007), studies in this area are limited. Twins discordant for obesity have shown significantly different gut microbial species composition (Tims et al., 2013). Lower BMI was associated with increased primary fibre degraders, while higher BMI subjects displayed an abundant network of butyrate producers, as such it was suggested that a shift in fermentation patterns near the end of the colon may be influencing energy homeostasis, although this study did not investigate the dietary impact on the microbial species population.

An analysis on Finnish MZ twin pairs concordant (*n*=9) and discordant (*n*=11) for BMI revealed dietary energy intakes and constituents to be more influential on microbial populations than BMI group (Simoes et al., 2013). Specifically, stool bacterial counts were significantly influenced by intakes of energy, monounsaturated and *n*-3 and *n*-6 polyunsaturated fatty acids, and soluble fibre intakes. Furthermore, profiles of *Bacteroides* spp. were very similar in co-twins consuming the same amount of energy or saturated fats than twins discordant for intakes. These findings suggest that the obesity status *per se* is less influential on microbial species population and rather, it is the content of the diet (i.e. high energy and fat) mediating important changes, highlighting the essential need to incorporate dietary variables into obesity research. The role of host genetics on the microbiome composition is still unclear, though our group has recently shown that gut microbiome phenotypes ranged from 0% to 39% heritable (Goodrich et al., 2014a), further studies (which are ongoing) are needed to provide clarity.

Recent findings suggest that the microbial species signature has the potential to promote and aggravate an obesity phenotype and may interact synergistically through diet-by-microbiota

interactions. In a unique study, transplantation of fecal samples from US female lean and obese co-twins (3 DZ; 1 MZ) into germ-free mice revealed the obese compared to lean co-twin's microbiota significantly increased adiposity and metabolic derangements in mice (Ridaura et al., 2013). When co-housed with lean mice, obese mice adiposity and metabolic effects were reduced, suspected to be a result of Bacteroidetes translocation from lean to obese mice. When fed a diet based on the lower tertile of saturated fats and upper tertile of fruits and vegetables of the US National Health and Nutrition Examination Survey (NHANES), obese mice adiposity was sustained, but relieved when co-housed with lean. Moreover, when compared to co-housing with normal chow, the healthier diet aided the success of lean to obese mice bacterial invasion. When supplied with an unhealthy diet containing the upper tertile of saturated fats with the lower tertile of fruits and vegetables from the NHANES survey, both lean and obese mice presented with significantly increased adiposity, mitigating transmissible effects. These results show the potentially causal role of the microbiome composition and potential for its manipulation.

1.3 Applying omics to untangle the effects of diet on metabolic disease in twins

Obesity elevates risk of cardiometabolic diseases such as the metabolic syndrome and CVD. The metabolic syndrome is a cluster of metabolic abnormalities characterised by at least three of the five symptoms: high blood pressure, an unfavourable lipid profile specifically elevated triglycerides and reduced high density lipoprotein cholesterol, elevated waist circumference, and elevated fasting blood glucose. One of the primary risk factors for cardiometabolic disease is elevated visceral fat mass (St-Pierre et al., 2002, Naukkarinen et al., 2014), which is metabolically active fat that covers the internal abdominal organs.

In the previous sections I have discussed the independent applications of metabolomics and the microbiome to examining diet, primarily through studies on body weight discordant MZ twins. However the link between diet and metabolic outcomes such as the metabolic syndrome and related factors could potentially be further elucidated through collaborative applications of omics technologies. In one recent twin study (Bogl et al., 2016), using metabolomics in combination with waist circumference (an indicator of visceral fat mass) a strong genetic overlap between waist

circumference and a number of metabolites, most notably positive correlations between phenylalanine (rg = 0.40), glycoprotein (rg = 0.37), serum triglycerides (rg = 0.36) and BCAAs (rg = 0.30–0.40), and negative correlations between HDL particle diameter (rg = − 0.33) and HDL cholesterol (rg = − 0.30), supporting the usage of the MZ discordant twin model.

A single study employed multi-omic methods to examine the time-dependent impact of a Big Mac challenge (Bondia-Pons et al., 2014) on 394 serum metabolites and faecal microbiota in 16 MZ twin pairs discordant for BMI and 9 concordant non-obese MZ pairs (Bondia-Pons et al., 2014). Overall, intra-pair changes in metabolite levels were minimal following the Big Mac challenge. Intra-pair differences at baseline for two branched chain amino acid (BCAAs) were found to converge after 120 minutes following the Big Mac challenge, with lower BCAAs in the high weight co-twin compared to the low weight co-twin at baseline. Secondary bile acids glycocholic acid and glycolithocholic acid were significantly different at 120 min following the Big Mac challenge, with higher levels in lean co-twins compared to higher weight co-twins. An examination into *Bacteroides* spp. diversity did not find significant differences between weight-discordant twins, though overall *Bacteroides* spp. diversity was positively associated with postprandial changes in sugar- and microbiota-derived organic acids. Links to other clinical indicators of metabolic health were also assessed, in particular postprandial changes in glycine-conjugated ursodeoxycholic acid were negatively associated with liver fat content as well as MI, a marker of insulin sensitivity. This study showed overall that within-pair similarity was the primary determinant of the metabolic response to a dietary challenge, highlighting the important role of genetics and early life factors in metabolism.

1.4 Conclusion

The interconnectedness between each omic presented here is evident. A systems biological approach that incorporates multiple omics methods for phenotype definition for use in nutriomic studies is much needed (Norheim et al., 2012). The principal central to this approach is that the whole organism will provide a more accurate view than the sum of its parts, or rather, the entire system has definitive characteristics which will not be replicated by simply adding the effects (MacLellan et al., 2012). However, a primary setback for amalgamating these methods is the inability to define what is “normal” due to the significant inter-individual variability underlying

metabolic processes (Kussmann et al., 2006). Furthermore, it has been suggested that many omics vary relatively little over time within individuals but significantly between individuals compounded by factors such as age, BMI and gender (Eady et al., 2005). To aid in discerning these effects, longitudinal phenotypic information from twin registries will be a critical resource for molecular dietary studies of the future (van Dongen et al., 2012). The studies presented here fall short of this view of an integrated technology-driven approach and focus rather on omics in isolation as this is where the current research stands. To move forward twin cohorts need to make a collaborative effort, collecting extensive dietary data in conjunction with multiple omics, undertaking complex statistical analyses, while contributing findings to international proposed databases such as the Nutritional Phenotype database (van Ommen et al., 2010).

Identifying individuals susceptible to poor dietary habits and defining an (un)healthy phenotype are the overarching aims of nutrigenetic and nutrigenomic research. Twins have provided valuable evidence that many dietary intakes are influenced by genetics, validating further nutrigenetic research and future dietary counselling which targets this domain. Early studies of twins for nutriomic studies have primarily been for unhealthy phenotype definition and single omic, while being less inclusive of diet. However, future studies need to begin to incorporate dietary factors to fill a void in this research area. Although accurate dietary assessment methods are problematic (Tucker et al., 2013). As twins have shown, metabolites are significantly correlated with dietary constituents, and could in the future be used as surrogates, as well as other techniques (e.g. microbiome) (Lloyd et al., 2013, Vernocchi et al., 2012). Twin studies will remain an important and integral part of nutritional research now and in the future.

Chapter 2 Hypothesis, aims and tasks of the thesis

2.1 Hypothesis

Metabolites have the potential to act as useful surrogates of dietary intakes. Identification of metabolites in tissues that consistently associate with reported intakes through validation by the co-twin control method can be used to better establish and explore diet-disease relationships.

2.2 Aims

i) To identify and characterise metabolite associates of self-reported intakes of foods and dietary patterns.

ii) To create, validate and test the utility of food group metabolite scores as novel biomarkers of food group intakes.

iii) To identify metabolite marker(s) of gut microbiome diversity that are modulated by food intake, identify the gut microbiome profile of those markers and their relationship to metabolic disease.

iv) To create a visceral fat mass dietary risk score and characterise the score using a multi-omic approach.

Chapter 3 Materials and methods

In this chapter I provide an overview of the study population used, I briefly describe the food frequency questionnaire used to evaluate food intakes, the metabolomics and genotyping datasets and clinical measures used throughout the thesis. Moreover I introduce the statistical method for calculating heritability.

3.1 Subjects and phenotypes

3.1.1 Subjects

Subjects included in all analyses were twins enrolled on the TwinsUK registry, a sample of extensively phenotyped, mainly female, adult MZ and DZ twins from the UK (Moayyeri et al., 2013, Spector and Williams, 2006). Healthy Caucasian twins were recruited nation-wide primarily through media campaigns. Twin zygosity was determined by a validated questionnaire or through multiplex DNA fingerprinting (PE Applied Biosystems, Foster City, CA). Ethical approval has been obtained from St. Thomas' Hospital Research Ethics committee and all subjects have undergone informed consent. Specifically, my thesis included female and male twins, aged 18 to 80 years, who had completed at least one FFQ between 1994 and 2001 (**Appendix B Document 1**), in 2007, and in 2014 and 2015 (**Appendix B Document 2**).

3.1.2 Food frequency questionnaire

Twins completed a 131-item FFQ that was developed and validated against pre-established nutrient biomarkers for the European Prospective Investigation into Diet and Cancer (EPIC) Norfolk (**Appendix B Documents 1 and 2**) (Bingham et al., 2001, Bingham et al., 1997). Processing of the FFQ, including subject exclusion and determination of nutrient intakes was completed by nutritionists from the University of East Anglia. Intake frequency of an average serving of listed foods was determined from a 9-point scale ranging from "Never or less than once/month" to "6+ per day". The questionnaire was intended to capture average intakes in the past year. Nutrient intakes were determined via consultation with *McCance and Widdowson's The Composition of Foods* 6th

edition (McCance et al., 1991), this resource includes nutritional information (including fat, sodium, fibre, and carbohydrates, vitamins and minerals) for over 1,200 commonly consumed foods derived from the UK Nutrient Databank.

3.1.2.1 Exclusion criteria

Submitted FFQs were excluded if greater than 10 food items were left unanswered, or if the total energy intake estimate derived from FFQ as a ratio of the subject's estimated basal metabolic rate (determined by the Harris-Benedict equation (Frankenfield et al., 1998)) was more than two standard deviations outside the mean of this ratio (< 0.52 or > 2.58). Overall, this included 828 twins (9.4%) and 284 twins (6.5%) who completed FFQs between 1994 and 2007 (*n*=8785) and between 2014 and 2015 (*n*=4400), respectively.

3.1.2.2 Residual energy adjustment

Differences in reported intakes as a product of physiological requirements (due to differences in body size, physical activity, and metabolic efficiency) may mask true diet-outcome associations. To account for this, food and nutrient intakes were energy adjusted using the residual method prior to analysis (Willett and Stampfer, 1986), unless otherwise specified. This involved regressing each food item or nutrient on the estimated total kilocalorie intake. The residual was then added to a constant, in this case, predicted food or nutrient intake for the mean estimated total kilocalorie intake.

3.1.3 Anthropometry

Anthropometric measures including body height (metres) and weight (kilograms) were taken at clinical visits by a trained research assistant. Body mass index (BMI) was calculated as body weight in kilograms divided by the square of height in metres.

3.1.4 Samples

3.1.4.1 Blood samples

Fasted blood samples were collected by a trained research nurse at the twin's annual visit at St. Thomas' Hospital. Twins were instructed to fast for 8 hours prior to their clinical visit. Visits occurred in the morning (9:00 or 10:00) or afternoon (13:00 or 14:00). Time of collection was not recorded

therefore diurnal variation may have influenced the results. Following collection samples were stored at -80°C until further processing.

3.1.4.2 Faecal samples

Faecal samples were collected by the twins at home. Following collection, samples were stored in the refrigerator for 2 days or less prior to their annual clinical visit at St. Thomas' Hospital. Once the samples arrived with the clinical team they were stored at −80°C until further processing.

3.1.5 Clinical measures

3.1.5.1 Blood pressure

Systolic and diastolic blood pressure (SBP and DBP, respectively) were carried out by an experienced nurse using the Marshall mb02 or the Omron Mx3 Digital Blood Pressure Monitors. An average of three readings (separated by one minute) was used for analysis.

3.1.5.2 Blood lipid profiling and glucose

Three devices (Cobas Fara; Roche Diagnostics, Lewes, UK; Kodak Ektachem dry chemistry analysers [Johnson and Johnson Vitros Ektachem machine, Beckman LX20 analysers, Roche P800 modular system]) were used to measure serum levels of total cholesterol, high density lipoprotein (HDL) cholesterol and triglycerides (TG) on fasted blood samples using enzymatic colorimetric assays. Using an enzymatic colorimetric slide assay fasting blood glucose was measured on an Ektachem 700 multichannel analyser (Johnson and Johnson Clinical Diagnostic Systems, Amersham, UK).

3.1.6 Metabolomics

3.1.6.1 Non-targeted metabolomics

Metabolites were detected and quantified in fasted serum and plasma samples by the metabolomics platform Metabolon, Inc. (Durham, NC), which uses a non-targeted mass spectrometry based-approach as has been previously described (Menni et al., 2013b) and described in greater detail below. Metabolomics was completed cross-sectionally in three batches, one in serum and two in plasma. Quality control of the non-targeted metabolomics dataset was undertaken by Metabolon Inc. and Dr Cristina Menni. Raw data were median-normalised by dividing

metabolite concentrations by the day median of that metabolite. As the metabolite concentrations were not normally distributed, all metabolites underwent a rank-based inverse normal transformation (INT). The rank-based INT method used back transformation of the sample quantile to estimate the expected normal scores (Beasley et al., 2009). Overall there were 456 metabolites, including 292 chemically identified ('known') and 164 for which the identity is currently unknown ('unknown'). In 2016 non-targeted metabolomics profiling was also completed on blood samples at 3 time points on 2069 twins, the same quality control procedures were undertaken, the details of this dataset are not included as only one metabolite was used for analysis in Chapter 6.

3.1.6.1.1 Metabolon metabolomics detailed methods

Methanol was added to samples for protein precipitation and isolation of a range of metabolites, samples were shaken vigorously for 2 minutes, centrifuged, and the resultant extract divided into four portions: one to be analysed by ultra-high performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS; positive mode), one to be analysed by UPLC-MS/MS (negative mode), one to be analysed by gas chromatography–mass spectrometry (GC-MS), and one sample saved in case of incident. Three different controls were used in conjunction with sample analysis, these included: human plasma samples that have been analysed considerably by Metabolon, Inc. for replication; water samples for process blanks; and multiple standard solutions spiked into all experimental samples to assess equipment performance. Randomisation of samples and controls was ensured across runs.

The UPLC-MS/MS platform used a Waters Acquity UPLC and a ThermoFisher LTQ mass spectrometer, the latter comprised of an electrospray ionization source and a linear ion-trap mass analyser. The instrument had scanning set at 99-1000 m/z and to fluctuate between MS and MS/MS. Positive and negative ions were identified in acidic and basic extracts, respectively by separate injections. Columns (Waters UPLC BEH C18-2.1 × 100 mm, 1.7 µm) had sample extracts loaded and then gradient-eluted with water and 95% methanol with 0.1% formic acid or 6.5 mM ammonium bicarbonate for acidic and basic extracts, respectively. Following each injection, columns were washed.

Analysis of samples by GC-MS were first vacuum dried for 18 h minimum before derivitization using bistimethyl-silyl trifluoroacetamide under nitrogen. Following derivitization,

separation of samples occurred on a 5% phenyldimethyl silicone column. The column used helium as the carrier gas at a temperature rise from 60° to 340° C for a run time of 17 minutes. Sample analysis was conducted on a Thermo-Finnigan Trace DSQ MS set at unit mass resolving power with electron impact ionization and an atomic mass unit scan range from 50-750.

Software developed at Metabolon was used for metabolite identification, which compared ion features (retention time, molecular weight (m/z), preferred adducts, MS spectra, etc.) in the experimental samples to a reference library (Dehaven et al., 2010). Known metabolites were identified by comparison to a mass spectroscopy library containing >2,400 purified standards. Approximately 5,300 unknown, but commonly reoccurring biochemicals have been added to this library for identification in the experimental samples. Peaks were quantified by area under the curve and therefore use relative measures and hence have no units.

3.1.6.2 Targeted metabolomics

A targeted metabolomic assay using the Biocrates Absolute IDQTM-kit p150 (BIOCRATES Life Sciences, AG, Innsbruck, Austria) was conducted on serum samples (by Biocrates employees) from 1030 twins who also had non-targeted metabolomics profiling from the same visit the details have been previously described (Illig et al., 2010, Römisch-Margl et al., 2012). In short, the flow injection analysis (FIA) tandem mass spectrometry (MS/MS) method quantifies 163 small molecule metabolites at once through multiple reaction monitoring. Metabolite quantification is completed by reference to the applicable internal standards. Concentrations of targeted metabolites are reported in µM.

The Biocrates platform analyses acylcarnitines (Cx:y), hydroxylacylcarnitines [C(OH)x:y] and dicarboxylacylcarnitines (Cx:y-DC); amino acids; sphingomyelins (SMx:y) and sphingomyelin-derivatives [SM(OH)x:y]; and glycerophospholipids (PC) in absolute metabolite values (mM). The metabolite concentrations were right-skewed and therefore log transformation was undertaken. There were 18 metabolites measured both by the Biocrates and Metabolon platforms, therefore for analyses including both platforms only 145 metabolites were considered from the Biocrates platform. **Table 3-1** shows the Pearson's correlation between metabolites measured on both platforms.

Table 3-1. Pearson's correlation between metabolites measured by the targeted and non-targeted platforms (*n*=1030)

Metabolite name	Pearson correlation (P<0.001)
Glutamine	0.3121
Tryptophan	0.5094
Histidine	0.3099
Phenylalanine	0.4804
Threonine	0.5537
Tyrosine	0.6411
Methionine	0.5298
Ornithine	0.1522
Valine	0.6377
Proline	0.8113
Acetylcarnitine	0.6423
Serine	0.4613
Hexanoylcarnitine	0.6972
Glycine	0.6759
Butyrylcarnitine	0.6375
Propionylcarnitine	0.6719
Octanoylcarnitine	0.7703
Decanoylcarnitine	0.7743

3.1.7 Genotyping

Genotyping for TwinsUK samples has been outlined recently (Metrustry et al., 2014). Quality control of the genotyping data was undertaken by Dr Massimo Mangino. Multiple Illumina arrays were used for genotyping: Human-Hap300 (Richards et al., 2008), HumanHap610Q, 1M-Duo and 1.2M-Duo 1M. First, normalized intensity data from each of the arrays were pooled independently (1M-Duo and 1.2M-Duo 1 M pooled together), then genotypes assigned for each dataset by the Illuminus calling algorithm (Teo et al., 2007). Had an individual's expected genotype been called with lower than a posterior probability cutoff of 0.95, no calls were assigned. Pooling was validated by a visual inspection for observable batch effects of 100 random, shared SNPs. SNPs were excluded in a similar manner from each of the three datasets by visually inspecting intensity cluster plots of significant SNPs for no calling bias due to overdispersion, and/or false assignment of genotype. Samples and SNPs were further excluded for the reasons outlined in **Table 3-2** (columns **A** and **B**, respectively). Each of the three datasets had alleles aligned to HapMap2 or HapMap3 forward strand alleles. A pair-wise comparison was then conducted between the three datasets that allowed

for additional sample and SNP exclusion for subsequent avoidance of false genotyping, as outlined in **Table 3-2** (column **C**). Finally, the three datasets were merged, reserving data from the array that had typed the largest number of SNPs when two different arrays were used for one individual.

Table 3-2. Exclusion criteria for samples, SNPs and following pair-wise comparison

A. Samples	**B.** SNPs	**C.** Pair-wise comparison
1. Sample call rate <98%	1. Hardy-Weinberg p-value $<10^{-6}$ (determined in a sample of unrelated subjects)	1. Similarity at corresponding samples <1%
2. Heterozygosity across all SNPs ≥ 2 SD from the sample mean	2. MAF <1% (determined in a sample of unrelated subjects)	2. Similarity at corresponding SNPs <1%
3. Identification of non-European ancestry by comparison of the PCA with HapMap3 populations.	3. SNP call rate < 97% for SNPs with MAF ≥ 5%, or <99% for 1% ≤ MAF < 5%	3. All pair-wise dataset comparisons inspected for logistic regression by visual investigation of QQ plots
4. Pair-wise IBD probabilities imply sample misidentification		4. Hardy-Weinberg p-value $<10^{-6}$ (determined in a sample of unrelated subjects)
5. Misidentified MZ and DZ twins were amended in accordance with IBD probabilities		5. Pair-wise IBD probabilities imply sample misidentification

Abbreviations: SNPs, single nucleotide polymorphisms; SD, standard deviation; PCA, Principal Components Analysis; IBD, identity-by-descendant; MZ, monozygotic; DZ, dizygotic; MAF, minor allele frequency; Quintile-Quintile, QQ.

3.2 Statistical analysis

3.2.1 Heritability

Heritabilities of dietary phenotypes were determined using linear structural equation modeling in Mx (Neale et al., 2003, Neale et al., 1992). The variance of the phenotype is decomposed into 3 different effect components: A, additive genetic; C, common environmental; E, non-shared environmental. This is termed univariate ACE modelling. If MZ twins are significantly more alike in a trait than DZ twins, it is indicated that additive genetic effects are high. The common environmental component indicates the effect of the family environment on a trait. In the model the effect of the family is assumed to be equal in both MZ and DZ twin pairs (Kyvik, 2000). The effects that are unique to each individual and also the measurement error are captured by the non-shared environmental component. Heritability (h^2) is the proportion of the variance in a trait due to genetic factors and is demonstrated by the equation, $h^2 = (A)/(A + C + E)$. Each of the four models: ACE, AE, CE, and E were tested and evaluated by the Akaike's information criterion (AIC); the lowest AIC

indicated the best fitting model reflecting a good balance between goodness of fit and parsimony (Neale et al., 1992). Prior to modelling, the phenotypes were adjusted for age and sex using linear regression and the residuals used in the subsequent analysis.

Chapter 4 Metabolomic associations of self reported food group intakes and food patterns

In this chapter I have undertaken a discovery approach, examining all metabolite associations with food intakes and dietary patterns in the TwinsUK population with the primary aim of identifying candidate food intake biomarkers. I applied the discordant MZ twin model as a novel method of replicating significant findings from the discovery analysis.

Part of this work has been published in *PLoS One* (Pallister et al., 2016).

4.1 Introduction

For decades epidemiological studies have depended on self-reported accounts of food intake to evaluate dietary intakes. Using these approaches, however it is difficult to precisely capture food intakes and portion sizes and consequently the findings of diet-disease association studies have been called into question. Using objective biomarkers of food intake (such as essential fatty acids in blood to estimate fatty fish intake), there is potential to improve upon this issue, but not many biomarkers have so far been identified and have scarcely been utilized in epidemiological settings.

The high-throughput profiling of metabolites completed through targeted and non-targeted means in tissues and biological fluids used in collaboration with dietary studies is deemed nutritional metabolomics. Diet and the chemicals within our bodies (metabolites) are intimately linked, as the food we consume provides the basic chemical inputs which our bodies use to produce metabolic products downstream, in addition to the energy required to complete these mechanisms that are essential to life.

Recently, metabolomics studies using non-targeted approaches have identified novel biomarkers of food intake in large population settings. Among those, subjects from the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial were used to identify 39 candidate dietary biomarkers for multiple food groups (Guertin et al., 2014b). In two separate studies of subjects from the African Americans in the Atherosclerosis Risk in Communities Study 39 metabolites were found to associate with alcohol intake (Zheng et al., 2014) and 48 to reported food intakes (Zheng et al.,

2014). Studies that have applied targeted metabolomics methods have also been successful in identifying significant associations between self-reported dietary intake patterns and serum metabolites (O'Sullivan et al., 2011, Altmaier et al., 2011, Floegel et al., 2013b).

A primary issue to be addressed is how candidate biomarkers of a healthy diet can be most effectively identified in collaboration with currently utilised methods of dietary assessment. Emerging research on dietary patterns is showing promising results within classic nutritional epidemiological studies (Hu, 2002). Empirically-derived dietary patterns are particularly relevant as they represent real trends in population reporting. Conversely, an advantage of using *a priori* over data-driven methods to detect metabolite biomarkers is that these patterns can be replicated in other populations. Due to its emerging importance for ameliorating derangements in metabolism associated with the metabolic syndrome (Kastorini et al., 2011), and the tendency of subjects in this group to underreport intakes (Kastorini et al., 2011), objective biomarkers of the Mediterranean diet would be particularly useful for clinicians and nutritionists.

Genetic factors have a strong impact on metabolism and may influence up to 81% of the variation in blood levels (Shin et al., 2014). As shown by classic examples of gene mutations causing inborn errors of metabolism that need dietary adjustments (e.g. phenylketonuria, maple syrup urine disease), complicated interactions exist between genes, metabolism and diet. However, recent studies suggest variation at a multiple loci with less exaggerated independent effects on metabolism are likely primarily responsible for the interplay between diet and its link to complex diseases (Kettunen et al., 2012). The effect of lipid metabolism genes (e.g. hepatic lipase gene, cholesteryl ester transfer protein) on the variation in blood cholesterol levels has been shown to depend on dietary fat content in recent intervention studies (Qi et al., 2015, Xu et al., 2015), however effects were minimal. Using more than 400 blood metabolites in the TwinsUK and the Cooperative health research in the Region of Augsburg (KORA) datasets (Wichmann et al., 2005, Shin et al., 2014), 145 metabolic loci were shown to associate with blood metabolites. The identified loci from these studies may pinpoint areas where metabolism, diet and genetics interact.

It may be challenging to replicate findings between populations due to the large inter-individual variability in metabolite levels (Sampson et al., 2013), to factors such as age (Menni et al., 2013b) and genetics (Shin et al., 2014). Identical (monozygotic, MZ) twins who are matched for age, sex and the baseline genetic sequence, may aid reproducibility issues by acting as “co-twin

controls". The TwinsUK group has used this method previously in one dietary metabolomics study (Menni et al., 2013c), where FFQ-derived principal component dietary patterns were found to correlated with 42 metabolites in the greater twin population with top findings replicated in MZ twins discordant for the dietary patterns.

Using blood samples profiled by both targeted and non-targeted metabolomic platforms, my aims for this chapter were to:

i) To identify novel associations with blood metabolites with reported food intakes and dietary patterns.
ii) To replicate these associations using MZ twins discordant for food intakes.
iii) To identify potential genotypes influencing food intakes from SNPs previously shown to be associated with metabolite levels on the TwinsUK and KORA cohorts (Shin et al., 2014).
iv) To provide the results of this analysis online through designing the DietMetab tool.

4.2 Materials and methods

I included 3559 female twins, who completed a 131-item validated FFQ between 1995 and 2007, and had Metabolon metabolomics and BMI data available within and including (+/-) 5 years of FFQ completion. I also analysed a subset of this sample who had Biocrates targeted metabolomics data (*n*=858).

The characteristics of the study population can be found in **Appendix C Table 1**.

4.2.1 Food items

Previously, intake frequencies were adjusted for total energy intake using the residual method (Willett and Stampfer, 1986) and I summed these into 71 food groups based on nutrient content, taste and usage (**Appendix C Table 2**).

4.2.2 Dietary patterns

4.2.2.1 Principal components

The principal components analysis (PCA) had been completed and the steps outlined previously (Teucher et al., 2007). The PCA was undertaken by nutritionists at the University of East Anglia. Briefly, the 131 food items were combined *a priori* into 54 groups based on nutrient content and culinary use. Intake frequencies of these groups were the sum of the servings per week for the original questionnaire items they were derived from. The residual energy adjusted food group intake frequencies (servings/week) were standardised to z scores and then used in the PCA. Bootstrapping is a resampling approach used to approximate properties of an estimator by evaluating those properties when sampling from an approximating distribution. It was used to test if family relatedness influenced the results. Briefly, the sampling distribution of the components was estimated in unrelated individuals. 10,000 bootstrap replications of the PCA were undertaken on subsamples containing one randomly selected twin. The full dataset PCA loadings were assessed against a 95% confidence interval determined from the 2.5% and 97.5% percentiles of this distribution. It was determined that family relatedness had no significant effect. The top 20 components were examined on a scree plot (**Appendix C Figure 1**), elbows were shown after components 1 and 5, however, components 2 to 5 were reasonable dietary patterns therefore the first 5 components were used. These 5 components, that explained 22% of the total variance, were denoted according to their top factor loadings

(**Table 4-1**) as follows: 'Fruit and vegetable'; 'High alcohol'; 'Traditional English'; 'Dieting'; and 'Low meat'. Consult **Appendix C Table 3** for their complete factor loadings.

Table 4-1. Principal component derived dietary patterns, with percentage of variance explained, and foods consumed in high and low intakes (Teucher et al., 2007)

Diet pattern	Variance explained	High intakes[1]	Low intakes[2]
Fruit & vegetable	8.2%	Fruit, allium and cruciferous vegetables	Fried potatoes
High alcohol	3.9%	Beer, wine and allium vegetables	High fiber breakfast cereals and fruit
Traditional English	3.6%	Fried fish and potatoes, meats, savoury pies and cruciferous vegetables	
Dieting	3.3%	Low-fat dairy products, low-sugar soda	Butter and sweet baked products
Low meat	3.2%	Baked beans, pizza and soy foods	Meat, other fish and seafood, and poultry

[1]Food frequency questionnaire items with factor loadings ≥0.20.
[2]Food frequency questionnaire items with factor loadings ≤-0.20.

4.2.2.2 Mediterranean diet score

I calculated Mediterranean diet adherence using the modified Mediterranean diet score (MDS) method, as outlined by Trichopoulou (Trichopoulou et al., 2005). Nine food/nutrient categories were included to generate a Mediterranean diet score (MDS) of 0 to 9 ("least" to "most" Mediterranean). The two nutrient intake categories included alcohol and fatty acid ratio (monounsaturated+polyunsaturated/saturated fatty acid intakes), both in grams per day and residual energy adjusted. Seven food group intakes were generated including: fruits (including nuts), vegetables, meats, fish, dairy, cereals, and legumes. To form groups, I first multiplied intake frequencies for assigned foods by the amount in grams per serving (**Appendix C Table 4**) and then divided by 7 to determine intake of that food in grams per day (raw data are per week). Next, the amount in grams of these foods was added to make the final category total. Categories in grams were then adjusted using the residual method (consult **Section 3.1.2.2**).

I then generated the scores. For all frequency categories as well as the fatty acid intake ratio, I determined median intakes of each category using all eligible twin data for FFQs 1 and 2 combined. I assigned a score of 0 (no MDS) or 1 (MDS) for each category depending on whether the twin was above or below the median intake, specific to the category (refer to **Table 4-2**). For alcohol intakes, a range was used for score assignment: twins between 5 and 25 g/d were assigned a score of 1, while those above or below this range were assigned a score of 0.

Finally, the MDS was then generated by summation of each category score to generate a score of 0 to 9.

Table 4-2. Criteria for a score assignment of 1 on the Mediterranean diet

Above median	Below median	Moderate
Fruit & nuts	Meat	Alcohol: 5 – 25 grams/d
Vegetables	Dairy	
Legumes		
Fish		
Cereals		
Unsaturated : saturated fatty acid		

4.2.3 Replication population

One metabolite was particularly strongly associated with low fat milk intake in the TwinsUK sample, I therefore obtained data on milk intake from the Cooperative Health Research in the Region of Augsburg (KORA) study to replicate this association. The KORA study includes individuals from the region of Augsburg, Germany who are unrelated (Holle et al., 2005). As part of the fourth survey (S4) between 1999-2001 4261 persons aged 25-74 years were examined. For the replication analysis, 1593 individuals with serum metabolite levels of the top milk-associated metabolite in the TwinsUK sample were used. Dietary intake in the KORA study was determined by a validated FFQ on the same day as blood sampling (Winkler and Doring, 1998)

The questionnaire included 24 food items and asked subjects to recall their "average intake" out of the six frequency categories: almost daily, several times per week, about once a week, several times per month, once a month or less, never. For the replication analysis I used the item "milk intake including buttermilk". Prior to analysis, I recoded the milk intake variable so interpretation of the result was easier therefore the least frequency ("never") was coded as 1 and the highest frequency ("almost daily") was coded as 6, and I changed the rest of the categories as follows: once a month or less, 2; several times per month, 3; about once a week, 4; and several times per week, 5.

4.2.4 Statistical analysis

Statistical analysis was carried out using Stata version 12.

For each metabolite, I ran a random intercept linear regression analysis using each food group as the predictor in the discovery sample, excluding discordant MZ twin pairs (pairs ≥ 1 SD apart in intake by food group). I adjusted for age, metabolite batch effects, BMI and family relatedness: $\Upsilon_i = \beta_0 + \beta_i X_{ij} + \gamma_i age_{ij} + \delta_i BMI_{ij} + \zeta_j + \varepsilon_{ij}$

where Y_i is the metabolite and X_{ij} is the food group intake of twin *j* from pair *i*, and ζ_j is the family-specific error component that captures the unobserved heterogeneity or family characteristics.

I used the Bonferroni correction to adjust for multiple testing which allowed a significant threshold of 1.08×10^{-6} (0.05/[77 diet phenotypes x 601 detected metabolites]). For associations that passed the significance threshold in the discovery sample, I repeated/replicated them using the same model in the MZ discordant twins sample. Associations were considered replicated in the MZ discordant sample if they (i) passed the 5% level of significance threshold and (ii) had the same effect direction as the discovery group (I used only this second criteria for the targeted platform). Lastly, using an inverse variance fixed effects meta-analysis I combined the results of both analyses, these are the results that I report here. To test if dietary patterns associated to a unique metabolomic signature not captured by investigating single food groups I repeated the analysis for each of the six diet patterns (five principal components and the MDS). The beta coefficients (β) I present for the significant associations represent the reported food amount in servings per week or dietary pattern score that corresponds to a 1 SD change in the metabolite level.

4.2.5 Diet pattern association with metabolite mediated by reported food intake

For each metabolite associated with both diet patterns and food group intakes first the proportion of the variance of metabolite was determined for each relevant diet pattern after taking into account all covariates (age, BMI, batch effects, family relatedness). This quantity is indicated as r^2_x. The proportion of the variance for the metabolite explained by the applicable diet pattern was then calculated after taking into account the same covariates as above but also including intakes of all associated foods (r^2_{xy}). The percentage of the diet pattern association mediated by the food group intake (r^2_y) was calculated as the proportion of the variance of metabolite that is due to the diet pattern association with the food intake, namely $1 - (r^2_{xy}/r^2_x)$.

4.2.6 Genotype associations

I have outlined the genotyping protocols that have been undertaken in the TwinsUK population in **Section 3.1.7**. A genome-wide association study (GWAS) was previously undertaken on the Metabolon metabolomics datasets conducted for TwinsUK and the Cooperative health research in the Region of Augsburg (KORA) cohorts (Wichmann et al., 2005, Shin et al., 2014). Based on the results of this study, I undertook diet-genotype associations on gene variants (56 SNPs) that were associated with blood levels of dietary-associated metabolites (48 metabolites). To identify

associations, I used additive genotype as a predictor of the appropriate energy-adjusted food group intake or dietary pattern, including age and family relatedness as covariates. I defined statistical significance using a Bonferroni correction of 2.94×10^{-4} (0.05/170 tests).

4.3 Results

4.3.1 Food group-metabolite associations

I found 178 significant associations with 39 food groups after meta-analysing the discovery and MZ discordant twin groups, consisting of 105 unique metabolites (**Appendix C Tables 5, 6, and 7**). From these 105 unique metabolites, 73 have been identified (**Figure 4-1**) and 32 are unknown at this time. Those 73 identified metabolites belong to six biochemical groups: 38 lipids, 16 amino acids, 14 xenobiotics, 3 carbohydrates, 1 cofactor/vitamin, and 1 peptide (**Figure 4-2a and b**). To my knowledge, 72 of the food associations with identified metabolites have not been previously identified in nutritional metabolomics using similar methods (**Appendix C Table 7**).

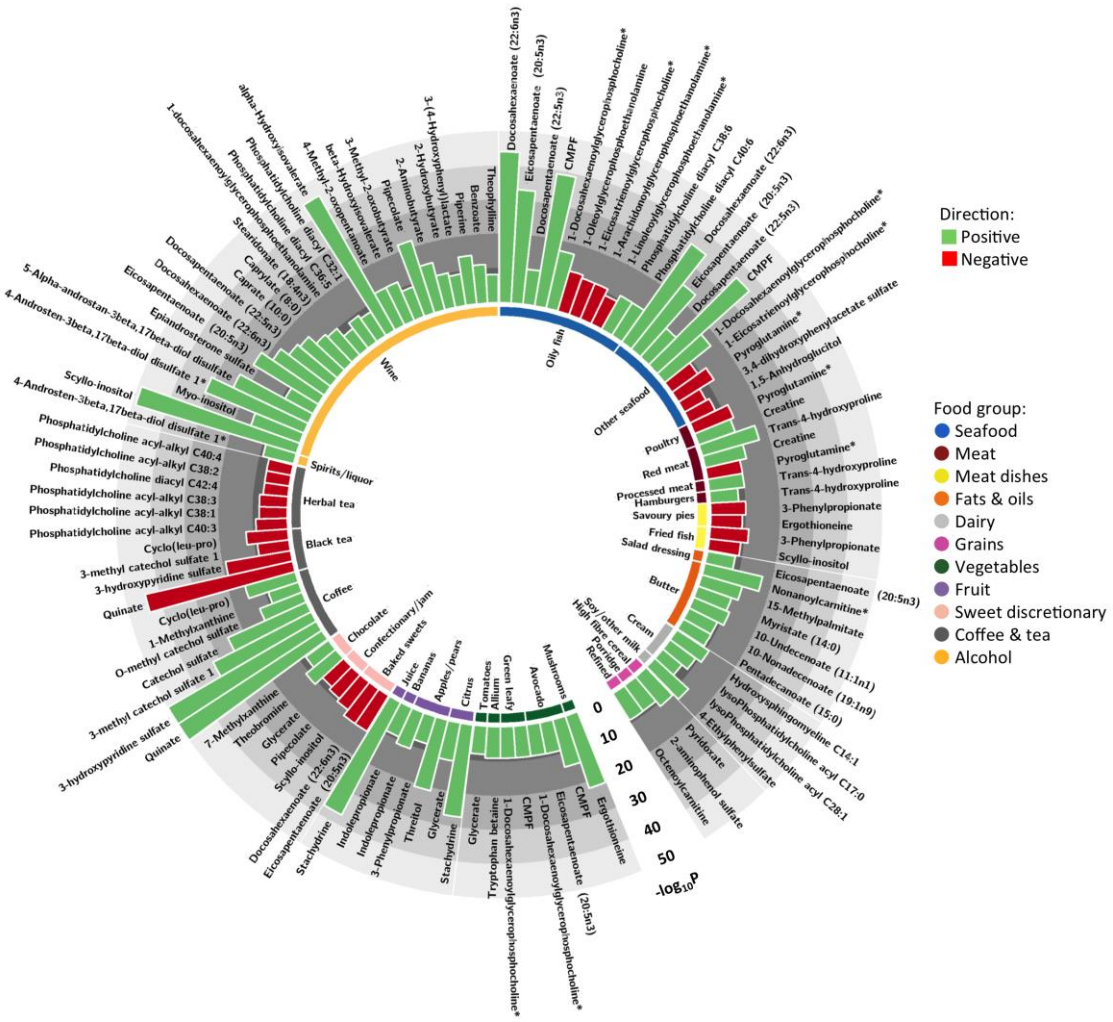


Figure 4-1. Associations between food group intakes and known metabolites
Associations between food group intakes and known blood metabolites are represented by the circular histogram plot. Associations are clustered according to general food groups, which are represented by the colored lines below the histogram bars. The histogram bars represent the $-\log_{10}$ of the p-value result from the fixed effects meta-analysis and the color of the bars indicates the direction of association: *green*, positive; *red*, negative.

Figure 4-2. Pathways represented by associated metabolites for general food groups

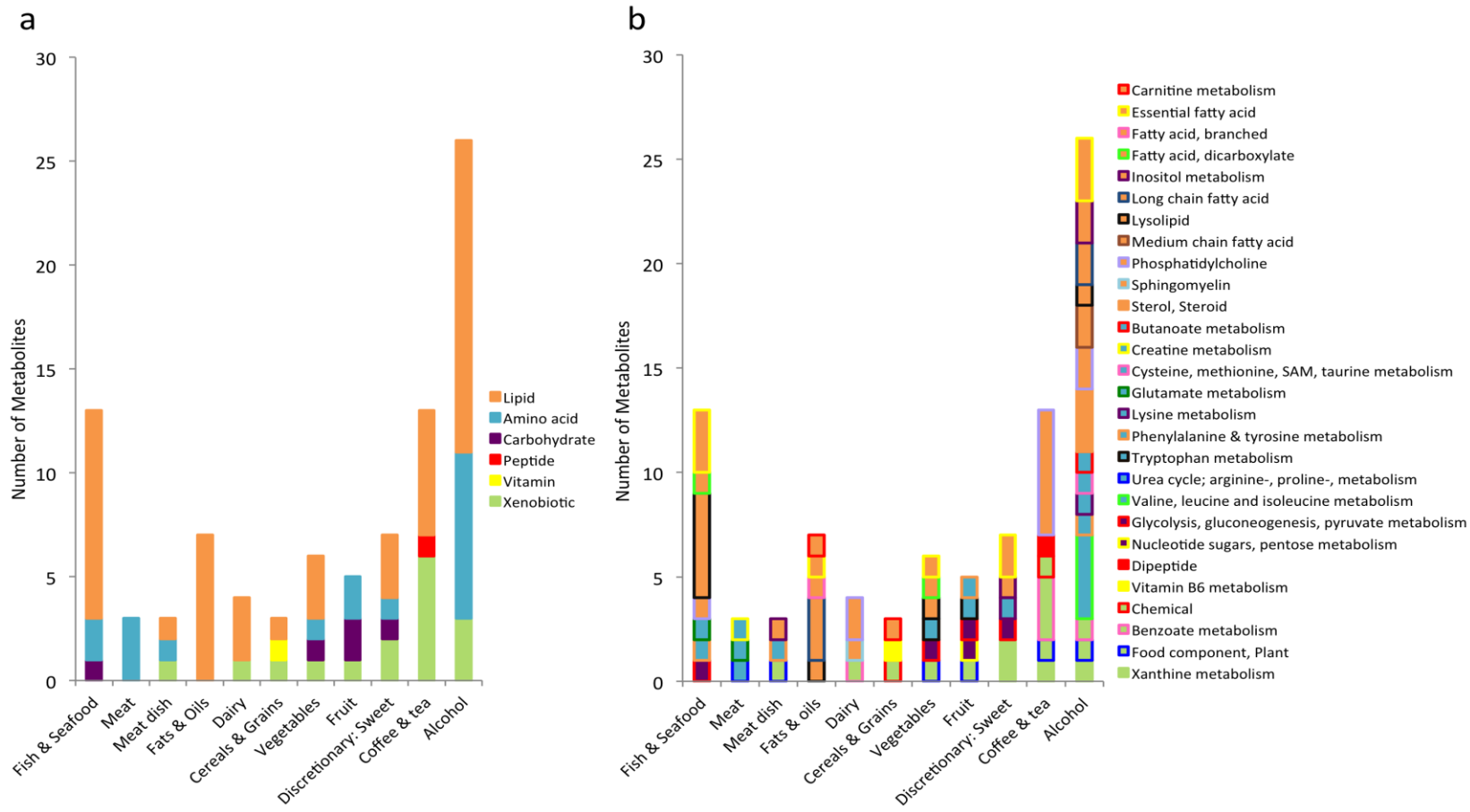


Figure 4-2 shows a stacked histogram of the number of associated metabolites representing each superpathway (a) and subpathway (b) by general food group intake. Pathways were determined by Metabolon based on Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Superpathways are assigned by chemical classes corresponding to the top-level KEGG pathways. Subpathways are assigned by the metabolic role of the metabolite.

I found reported intakes of alcoholic beverages to be associated with the most metabolites, including 41 associations overall (38 wine,1 spirits/liquors) with 15 associations (of the known metabolites) I believe to be novel; the top association was between wine and scyllo-inositol (meta-analysis result: Beta[SE]: 0.052[0.003]; P=1.47x10^{-49}). Reported intakes of teas and coffee included 27 associations (8 novel of the known metabolites) overall including: 12 coffee, 9 black tea, and 6 herbal tea. Notably, for both coffee and tea, 6 metabolites belonging to the xenobiotics pathway, i.e. metabolites foreign to the human body, were identified. These likely originated from gut bacterial metabolism of caffeine (e.g. catechol sulfate, 1-methylxanthine). The top association was between coffee and the unknown metabolite X-14473 (0.038[0.001]; P=6.12x10^{-187}). There were 26 associations (4 novel) with reported seafood consumption, including: 15 oily fish and 11 with other seafood. A number of the metabolites associated with reported seafood consumption were essential fatty acid metabolites; the top association was between docosahexaenoate (DHA; 22:6n3) and oily fish (0.177[0.013]; P=2.09x10^{-44}). I found 8 associations (5 novel) with meat products, specifically: 3 meat, 2 poultry, 2 processed meat and 1 beef burgers. The associations were primarily with amino acids, the top association being a novel associations between meat and trans-4-hydroxyproline (0.075[0.009]; P=1.08x10^{-17}). I found 14 metabolites (5 novel) were associated with reported dairy product intake, including: 9 butter, 3 cream and 2 low-fat milk. Primarily associations with dairy products included lipids, the top association was reported between low fat milk intake and the recently named metabolite trimethyl-N-aminovalerate (X-21365: 0.076[0.007]; P=9.36x10^{-27}). There were 10 associations (3 novel) with grain-rich foods, including: 5 high fibre breakfast cereals, 2 refined bread and grains, 2 porridge and 1 wholemeal bread and grains. The top association was between porridge and the unknown metabolite X-09789 (0.094[0.008]; P=4.96x10^{-33}). There were 13 associations (4 novel) with reported fruit intakes consumption, including: 6 apple and pears, 3 citrus fruits, 1 bananas, 1 berries, 1 peaches and 1 fruit juices. The top association within the fruit category was between fruit juice and stachydrine (0.058[0.005]; P=3.26x10^{-37}). I identified 16 associations (8 novel) with reported vegetable intakes, including: 6 green leafy, 5 avocado, 3 allium, 1 tomatoes and 1 mushrooms. The strongest association within the vegetables group was between reporeted mushroom consumption and ergothioneine (0.181 [0.019]; P=5.93x10^{-22}). I found 17 associations (9 novel) with reported sweet and savoury discretionary food intakes, including: 5 sweet baked products, 4 savory pies, 4 fried fish, 3 confectionary and jam, and 1 savoury snacks. Within the

discretionary foods group, the strongest association was between reported savoury snack intake and the unknown metabolite X-11372 (0.051[0.007]; P=3.88x10^{-14}). Other notable associations included 2 metabolites with reported chocolate intake (1 novel; top association with theobromine: 0.024[0.003]; P=1.34x10^{-11}), a novel association between reported soymilk intake and 4-ethylphenylsulfate (0.239[0.033]; P=6.05x10^{-13}) and also an association between an unknown metabolite and reported intake of soyfoods (X-11381: -0.108[0.020]; P=5.80x10^{-8}) and nuts (X-11315: 0.054[0.005]; P=3.75x10^{-25}).

A number of metabolites were associated to more than one food item, many of these are listed in **Tables 4-3** and **4-4**. Notably, the essential fatty acids eicosapentaenoate (EPA) and docosahexaenoate (DHA), their downstream metabolite 1-docosahexaenoylglycerophospho-choline and 3-carboxy-4-methyl-5-propyl-2-furanpropanoate were all positively associated with oily fish and other fish/seafood intakes, and all but DHA with avocado. These multiple associations may be due to correlated food intakes, though each source may be contributing independently. Higher levels of EPA and DHA were also associated with lower intakes of sweet baked products. The metabolite 3-phenylpropionate was associated positively to intakes of apples/pears and negatively with fried fish and savoury pies. The latter associations may be due to systematic under reporting rather than biological in origin as is discussed later. A further example of this observation are the metabolites quinate, cyclo(leu-pro), 3-methyl catechol sulfate 1 and 3-hydroxypyridine sulfate, which were positively associated with coffee intake and negatively with tea intake.

4.3.1.1 Milk association replication

I replicated the association between the previously unknown metabolite trimethyl-N-aminovalerate (X-21365) and milk intake in the KORA population (0.008[0.002]; P=6.88x10^{-6}), establishing the quality of the data.

4.3.2 Dietary pattern results

4.3.2.1 Nutrient profile of the Mediterranean diet

Energy-adjusted nutrient intakes are represented as percentages of the recommended intakes by tertile of the Mediterranean Diet Score (MDS) are presented in **Figure 4-3** with significance level for trend. Nutrient profiles of the FFQ PCs have been described previously (Teucher et al., 2007). There were significant trends for most nutrients from the bottom to top tertile of the MDS,

suggesting that individuals who scored higher on the MDS reported consuming more nutrient-dense diets, high in fibre and lower in total, saturated fat and sucrose content. Despite the monounsaturated fat content being lower in those who scored in the top versus bottom tertile of the MDS, an important characteristic of the Mediterranean diet derived from olive oil intake, a logistic regression analysis of olive oil intake (yes or no) was strongly predicted by the MDS (OR= 1.15, SE= 0.02, P=1.19x10^{-13}).

4.3.2.2 Heritability of the MDS

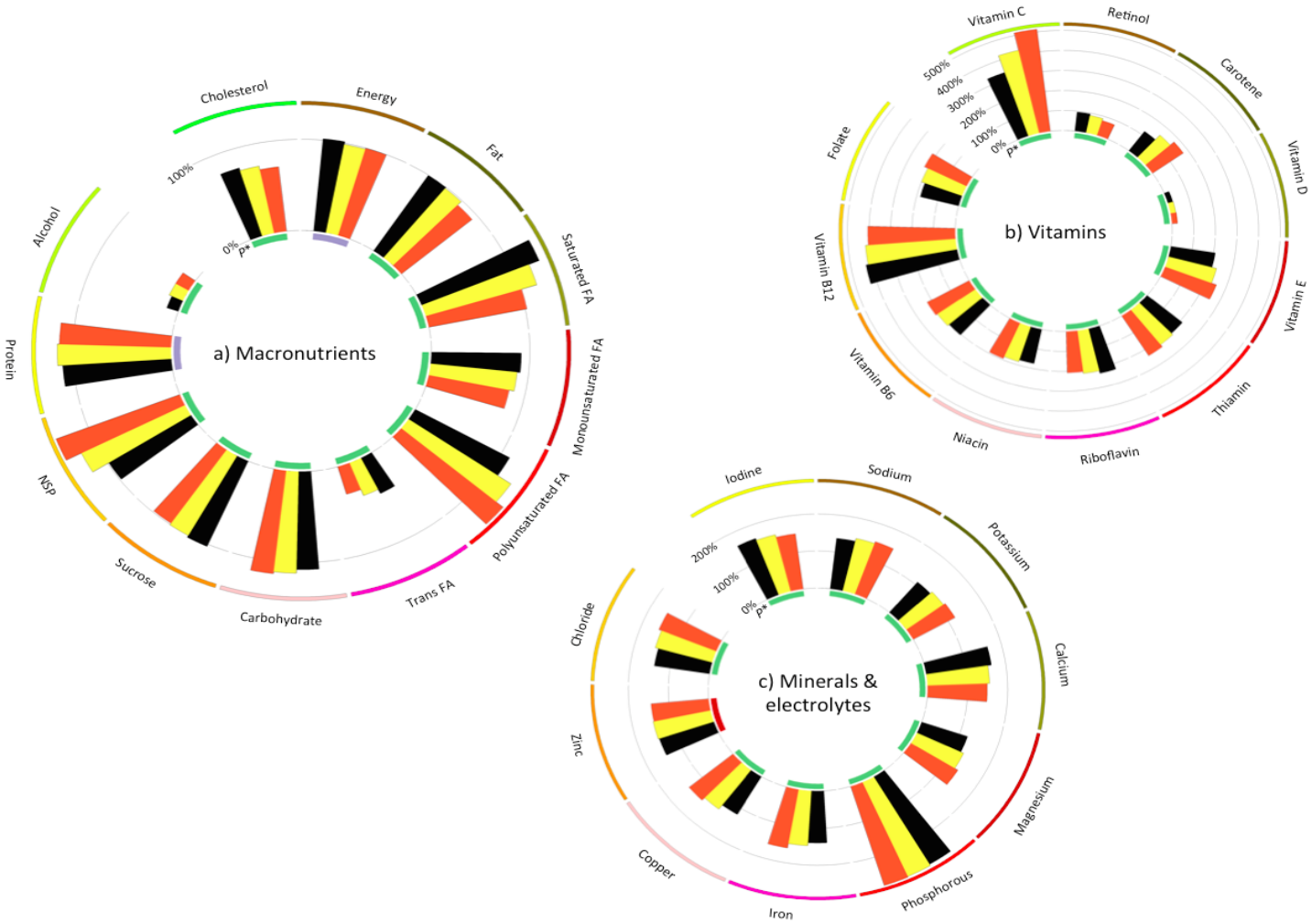
The best fitting model for MDS was the AE model, with heritability estimates of 0.41[95% CI = 0.36; 0.46] and 0.59[0.54;0.64] respectively.

4.3.2.1 Dietary pattern-metabolite associations

I identified 108 significant dietary pattern-metabolite associations that passed the Bonferroni level of significance (P<1.08x10^{-6}) in the discovery twin population and following meta-analysis (**Appendix C Tables 8, 9**). The MDS (**Figure 4-4a)** was associated with 18 metabolites (2 amino acids, 5 lipids, 2 xenobiotics and 9 unknown metabolites). Of the known metabolites, MDS was most strongly associated with DHA (meta-analysis result: 0.097[0.009]; P=2.45x10^{-27}). The F&V diet pattern (**Figure 4-4b**) was associated with 33 metabolites (6 amino acids, 3 carbohydrates, 1 cofactor/vitamin, 7 lipids, 1 peptides, 3 xenobiotics and 12 unnamed metabolites), a number of which were also associated with the MDS. A high F&V pattern was most strongly associated with glycerate (0.094[0.009]; P=1.97x10^{-27}), a carbohydrate involved in glycolysis, gluconeogenesis, and pyruvate metabolism. The high alcohol pattern (**Figure 4-4c**) was associated with 23 metabolites (3 amino acids, 1 cofactor/vitamin, 4 lipids, 3 xenobiotics, and 12 unnamed metabolites). The strongest known metabolite association with a high alcohol diet pattern was with the amino acid involved in valine, leucine and isoleucine metabolism, α-hydroxyisovalerate (0.158[0.013]; P=3.20x10^{-35}). The traditional English diet pattern (**Figure 4-4d)** was associated with 6 metabolites (2 amino acids, 1 xenobiotic, and 3 unnamed metabolites). A negative association was the strongest observed between the traditional English pattern and stachydrine (-0.094[0.012]; P=9.94x10^{-16}), an amino acid involved in urea cycle and arginine, proline metabolism. The dieting pattern (**Figure 4-4e)** was associated with 5 metabolites (2 xenobiotics, and 3 unnamed metabolites). Dieting was most strongly, negatively associated with quinate (-0.095[0.015]; P=1.40x10^{-10}), a xenobiotic derived from benzoate metabolism. The low meat pattern (**Figure 4-4f)** was associated with 22 metabolites (7 amino

acids, 7 lipids, and 8 unnamed metabolites). Most prominent, was a strong negative association with the furan fatty acid 3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF) (-0.144[0.015] P=3.63x10^{-21}).

Figure 4-3. Percentage of recommended intake of (a) macronutrients, (b) vitamins and (c) minerals and electrolytes by tertile of MDS



Average intakes by tertile (1, black; 2, yellow; 3, red) were assessed for percentage of recommended intakes. *P** is the significance value for the results of a linear regression using tertiles as predictors of the nutrient intakes (green, *P*<0.001; purple, *P*<0.05; red, not significant).

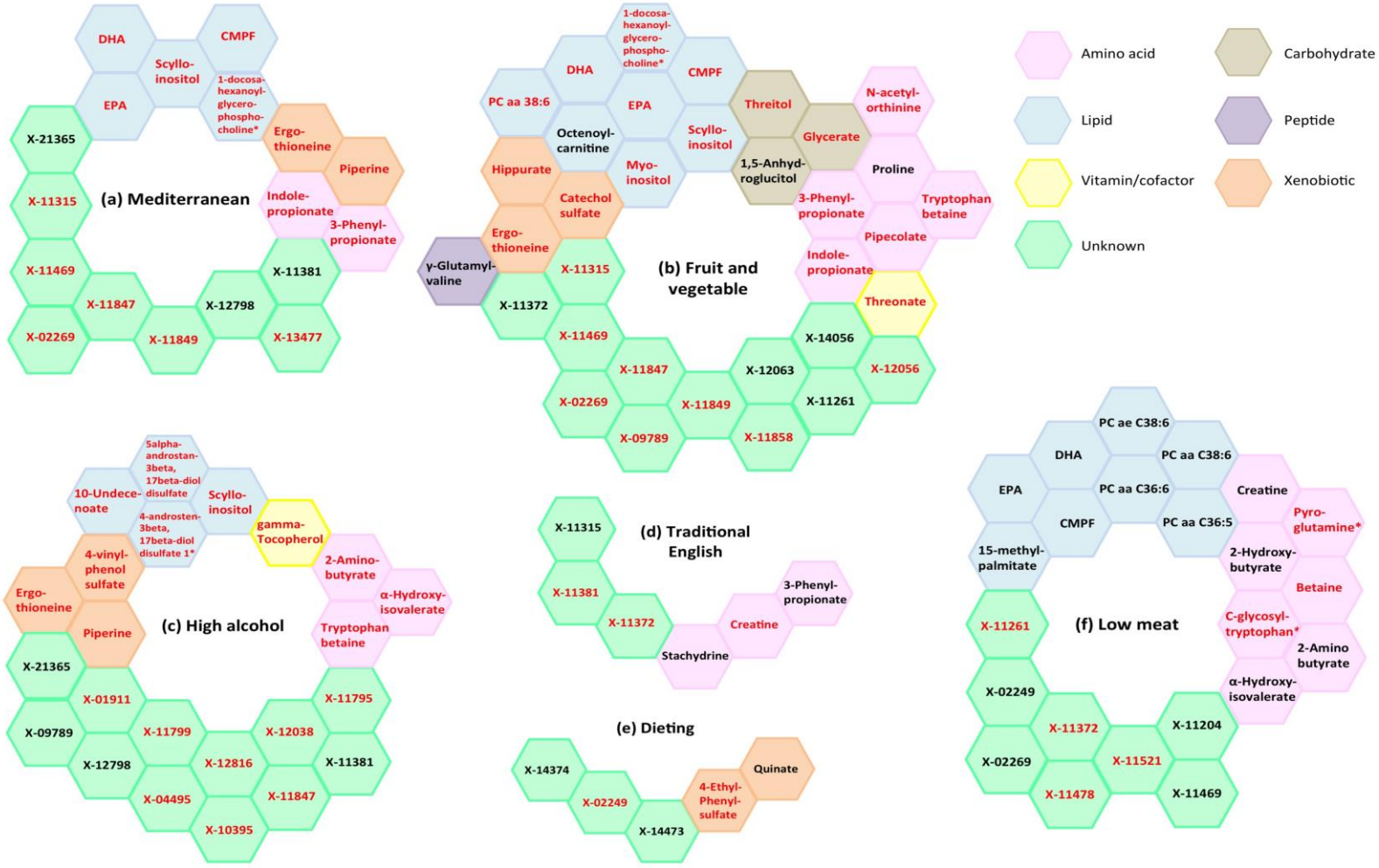


Figure 4-4. Significant diet phenotype and metabolite associations

Associations are significant at the Bonferroni level: 1.08×10^{-6} for discovery and meta-analysis. Significance level for discordant twin replication set ($P<0.05$). Names are colour coded according to direction of association: red, positive; black, negative.

4.3.2.2 Metabolites uniquely associated with food intakes or dietary patterns and those associated to both

Overall 126 metabolites (43 unknown) were associated to at least one dietary phenotype (**Figure 4-5**). Fifty-five metabolites were associated with single food intakes but not the dietary patterns (**Figure 4-5a**). This may suggest these metabolites are less important for dietary intakes of the whole population. Twenty-one metabolites were not associated to single foods (**Figure 4-5b**) but were associated with the dietary patterns. In this case the combined food intakes may have an effect on these metabolites not occurring by the food intake alone or the associations may originate from food intakes not explicitly measured by the FFQ or behaviours correlated to a dietary pattern. Fifty metabolites were associated to at least one pattern and one food (**Figure 4-5c**), there are likely the most important metabolites related to food intake in our dataset.

4.3.2.3 Dietary pattern-metabolite associations mediated by food intakes

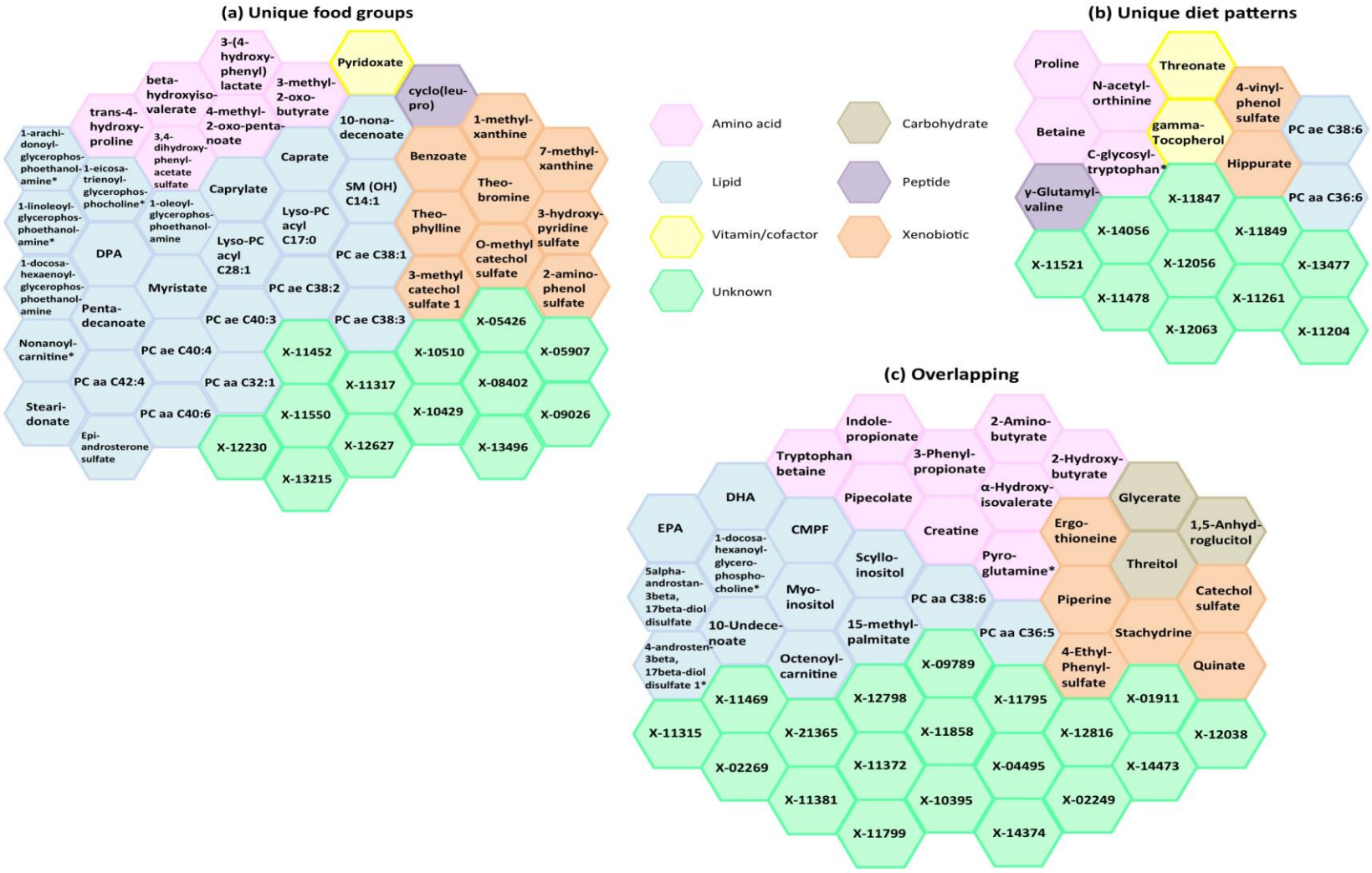
To confirm if the 50 metabolites associated with dietary patterns and at least one food group were driven by the reported food intakes and not other factors, I analysed the degree to which the variance in dietary pattern-metabolite associations was mediated by the metabolites association with single food intakes. I found that between <0% and 100% of the variance in the association between the metabolites and the dietary pattern scores were mediated by the metabolites association with food intake (**Tables 4-3** and **4-4**). The MDS and fruit and vegetable diet pattern associations with essential fatty acid metabolites (eicosapentaenoate, 1-docosahexaenoylglycerophosphocholine*, CMPF and docosahexaenoate) were strongly accounted for by food intakes (range: 86.2% to 99.7%). This was also the case for the low meat pattern association with pyroglutamine* which was strongly accounted for by meat intakes (93.2%), the dieting pattern with quinate which was accounted for by coffee and tea intakes (97.9%), the high alcohol pattern with 4-androsten-3beta, 17beta-diol disulfate 1*, 5alpha-androstan-3beta,17beta-diol disulfate and scyllo-inositol accounted for by wine or spirit intake (98.3%, 92.4% and 99.5%, respectively).

4.3.2.4 Gene-variants SNP associations

A GWAS study was previously performed on the TwinsUK dataset to identify SNPs associated with blood metabolites levels from the Metabolon platform (Shin et al., 2014). Of these 126

dietary-associated metabolites, 48 contributed to 62 metabolite-SNP associations including 56 unique SNPs in 39 genes (**Appendix C Table 10**). I undertook this analysis to identify potential SNPs which may be influencing dietary intakes. I did not find any SNP-diet associations with these variants at the Bonferroni level of statistical significance ($P<2.94x10^{-4}$ = 0.05/170 tests). I found 19 SNP-diet associations at the nominal level ($P<0.05$; **Appendix C Table 11**). The fact that I did not identify any significant associations between diet and metabolite-related SNPs suggests that these SNPs do not have such a strong effect as to influence eating behaviour, although it could also be due to the limited power due to the low sample size or the lack of precision by FFQs.

Figure 4-5. Metabolites significantly associated with only reported intakes of foods (a) and dietary patters (b), and to both (c)



Associations are significant at the Bonferroni level: 1.08×10^{-6} for discovery and meta-analysis. Significance level for discordant twin replication set: non-targeted = $P<0.05$; targeted = same direction.

Table 4-3. List of known metabolites associated with food intakes and dietary patterns and the degree to which their association with each pattern is through reported food intake

Metabolite	No. foods	Foods associated	R^2 all foods	Diet patterns associated	R^2 pattern no foods (1)	R^2 pattern with foods (2)	% association through food
eicosapentaenoate	6	↑ Avocado, High fat salad dressing, Oily fish, Other fish/seafood, Wine ↓ Sweet baked products	0.08	↑ MDS	0.026	0.001	95.1
				↑ F & V	0.047	0.004	90.9
				↓ Low Meat	0.025	0.004	86.1
1-docosahexaenoylglycerophospho-choline*	4	↑ Avocado, Green leafy vegetables, Oily fish, Other fish/seafood	0.04	↑ F & V	0.023	0.002	93.2
				↑ MDS	0.022	0.003	86.2
3-carboxy-4-methyl-5-propyl-2-furanpropanoate	4	↑ Avocado, Green leafy vegetables, Oily fish, Other fish/seafood	0.10	↑ F & V	0.029	0.000	99.7
				↑ MDS	0.028	0.001	97.5
				↓ Low Meat	0.036	0.015	59.2
docosahexaenoate (DHA; 22:6n3)	4	↑ Oily fish, Other fish/seafood, Wine ↓ Sweet baked products	0.11	↑ F & V	0.043	0.003	93.4
				↑ MDS	0.035	0.003	91.1
				↓ Low Meat	0.027	0.004	85.0
3-phenylpropionate (hydrocinnamate)	3	↑ Apples/pears ↓ Fried fish, Savoury pies	0.03	↑ F & V	0.027	0.006	76.8
				↓ Traditional English	0.015	0.006	62.0
glycerate	3	↑ Citrus fruit, Tomatoes Confectionary/jam	0.03	↑ F & V	0.037	0.009	75.8
pyroglutamine*	3	↓ Meat, Other fish/seafood, Poultry	0.04	↑ Low Meat	0.022	0.002	93.2
scyllo-inositol	3	↑ Wine ↓ Fried fish, Sweet baked products	0.13	↑ High Alcohol	0.043	0.000	99.5
				↑ F & V	0.035	0.010	71.3
				↑ MDS	0.015	0.010	30.6
4-androsten-3beta,17beta-diol disulfate 1*	2	↑ Spirits/liquor, Wine	0.09	↑ High Alcohol	0.035	0.001	98.3
creatine	2	↑ Meat, Poultry	0.03	↑ Traditional English	0.014	0.001	89.9
				↓ Low Meat	0.015	0.002	87.9
ergothioneine	2	↑ Mushrooms ↓ Savoury pies	0.12	↑ F & V	0.048	0.006	88.4
				↑ MDS	0.025	0.007	71.1
				↑ High Alcohol	0.041	0.022	47.0
indolepropionate	2	↑ Apples/pears, Bananas	0.03	↑ F & V	0.029	0.012	58.2
				↑ MDS	0.021	0.011	50.5
pipecolate	2	↑ Wine ↓ Confectionary/jam,	0.03	↑ F & V	0.016	0.007	58.5
quinate	2	↑ Coffee ↓ Black tea	0.20	↓ Dieting	0.014	0.000	97.9
stachydrine	2	↑ Citrus fruit, Fruit juice	0.11	↓ Traditional English	0.019	0.010	45.5
1,5-anhydroglucitol (1,5-AG)	1	↓ Other fish/seafood	0.01	↓ F & V	0.022	0.013	41.6
10-undecenoate (11:1n1)	1	↑ Butter	0.02	↑ High Alcohol	0.011	0.006	42.5
15-methylpalmitate (isobar with 2-methylpalmitate)	1	↑ Butter	0.03	↓ Low Meat	0.016	0.009	46.0
2-aminobutyrate	1	↑ Wine	0.02	↑ High Alcohol	0.017	0.003	83.1
				↓ Low Meat	0.018	0.008	57.4
2-hydroxybutyrate (AHB)	1	↑ Wine	0.02	↓ Low Meat	0.011	0.004	60.2
4-ethylphenylsulfate	1	↑ Soy/other milk	0.06	↑ Dieting	0.015	0.003	81.0
5alpha-androstan-3beta,17beta-diol disulfate	1	↑ Wine	0.06	↑ High Alcohol	0.029	0.002	92.4
alpha-hydroxyisovalerate	1	↑ Wine	0.07	↓ Low Meat	0.014	0.001	91.5
				↑ High Alcohol	0.049	0.008	84.1

Table 4-3. List of known metabolites associated with food intakes and dietary patterns and the degree to which their association with each pattern is through reported food intake

Metabolite	No. foods	Foods associated	R^2 all foods	Diet patterns associated	R^2 pattern no foods (1)	R^2 pattern with foods (2)	% association through food
catechol sulfate	1	↑ Coffee	0.03	↑ F & V	0.014	0.015	-3.5
myo-inositol	1	↑ Wine	0.02	↑ F & V	0.012	0.008	34.7
Octenoylcarnitine	1	↑ White/brown bread & refined grains	0.03	↓ F & V	0.038	0.019	50.3
Phosphatidylcholine diacyl C36:5	1	↑ Wine	0.05	↓ Low Meat	0.062	0.033	46.6
Phosphatidylcholine diacyl C38:6	1	↑ Oily fish	0.04	↑ F & V	0.031	0.008	73.4
				↓ Low Meat	0.036	0.027	25.8
piperine	1	↑ Wine	0.03	↑ High Alcohol	0.031	0.009	71.2
				↑ MDS	0.011	0.009	16.5
threitol	1	↑ Apples/pears	0.03	↑ F & V	0.016	0.005	71.2
tryptophan betaine	1	↑ Allium vegetables	0.03	↑ F & V	0.026	0.007	74.6
				↑ High Alcohol	0.028	0.011	62.0
				↑ MDS	0.024	0.010	57.3

F & V, fruit and vegetable pattern; MDS, Mediterranean diet score

(1) The proportion of the variance in the diet pattern explained by the metabolite after taking into account all covariates (age, BMI, batch effects and family relatedness).

(2) The proportion of the variance in the diet pattern explained by the metabolite after taking into account all covariates as in (1) and adjusting for all foods associated to the metabolite.

Table 4-4. List of unknown metabolites associated with both food intakes and dietary patterns and the degree to which their association with each pattern is mediated by reported food intakes

Metabolite	RT	Mass	No. foods	Foods associated	R^2 all foods	Diet patterns associated	R^2 pattern no foods (1)	R^2 pattern with foods (2)	% association through food
X-11315	1.19	130.2	13	↑ Apples/pears, Berries, Citrus fruit, Green leafy vegetables, High fibre breakfast cereals, Nuts, Oily fish, Peaches ↓ Confectionary/jam, Fried fish, Savoury pies, Sweet baked products, White/brown bread & refined grains	0.13	↑ F & V ↑ MDS ↓ Traditional English	0.088 0.042 0.019	0.001 0.002 0.004	98.4 94.6 79.3
X-02269	1.55	255.1	6	↑ Allium vegetables, Avocado, Green leafy vegetables, High fibre breakfast cereals, Oily fish, Other fish/seafood	0.11	↑ F & V ↑ MDS ↓ Low Meat	0.051 0.039 0.020	0.002 0.002 0.005	96.5 94.8 75.9
X-09789	2.62	153.1	5	↑ Apples/pears, High fibre breakfast cereals, Porridge, Wholemeal bread/grains	0.09	↑ F & V ↓ High Alcohol	0.016 0.026	0.001 0.004	98.2 90.1
X-11372	NA	NA	5	↑ Fried fish, Savoury pies, Savoury snacks ↓ Apples/pears, Green leafy vegetables	0.08	↓ F & V ↑ Traditional English ↑ Low Meat	0.046 0.015 0.016	0.004 0.005 0.014	91.6 65.1 14.4
X-11469	3.82	239.1	5	↑ Avocado, Green leafy vegetables, High fibre breakfast cereals, Oily fish, Other fish/seafood	0.11	↓ Low Meat ↑ F & V ↑ MDS	0.021 0.056 0.040	0.001 0.002 0.002	97.2 95.7 94.8
X-11381	1.11	186.2	2	↑ Processed meats ↓ Soy foods	0.03	↑ Traditional English ↓ MDS ↓ High Alcohol	0.011 0.017 0.012	0.003 0.011 0.013	76.6 36.4 -9.2
X-11799	1.58	226.0	2	↑ Wine ↓ Sweet baked products	0.06	↑ High Alcohol	0.030	0.001	96.0
X-12816	NA	NA	2	↑ Coffee ↓ Black tea	0.15	↑ High Alcohol	0.015	0.002	85.8
X-14374	NA	NA	2	↑ Coffee ↓ Black tea	0.09	↓ Dieting	0.010	0.000	98.0
X-14473	3.27	211.2	2	↑ Coffee ↓ Black tea	0.28	↓ Dieting	0.033	0.001	97.0
X-01911	4.26	464.1	1	↑ Wine	0.03	↑ High Alcohol	0.030	0.007	77.6
X-02249	4.03	267.2	1	↑ Butter	0.02	↑ Dieting ↓ Low Meat	0.013 0.019	0.004 0.011	71.8 40.0
X-04495	NA	NA	1	↑ Wine	0.02	↑ High Alcohol	0.016	0.001	93.8
X-10395	9.94	156.0	1	↑ Wine	0.03	↑ High Alcohol	0.017	0.001	91.7
X-11795	NA	NA	1	↑ Wine	0.06	↑ High Alcohol	0.043	0.007	83.4
X-11858	NA	NA	1	↑ Allium vegetables	0.03	↑ F & V	0.026	0.006	77.1
X-12038	5.82	245.3	1	↑ Wine	0.01	↑ High Alcohol	0.008	0.001	88.6
X-12798	1.84	240.1	1	↑ Low fat mik	0.02	↓ High Alcohol ↓ MDS	0.020 0.016	0.008 0.008	59.4 48.8
X-21365(3)	NA	NA	1	↑ Low fat mik	0.04	↓ High Alcohol ↓ MDS	0.017 0.019	0.005 0.009	73.7 54.1

RT, retention time; F & V, fruit and vegetable pattern; MDS, Mediterranean diet score; NA, not available

(1) The proportion of the variance in the diet pattern explained by the metabolite after taking into account all covariates (age, BMI, batch effects and family relatedness).

(2) The proportion of the variance in the diet pattern explained by the metabolite after taking into account all covariates as in (1) and adjusting for all foods associated to the metabolite.

(3) Metabolite recently identified as trimethyl-N-aminovalerate.

4.4 Discussion

Using one of the largest datasets of its kind, I identified and replicated using MZ discordant twins 73 novel associations between metabolites and specific food groups, providing further potential intake biomarkers for future research. Moreover, I found 108 significant metabolite associations with 5 different dietary patterns. Forty-eight metabolites contributed to 62 metabolite-SNP associations in 56 unique SNPs that were identified from a previous metabolomics GWAS study (Shin et al., 2014), though I did not find these SNPs to be related to dietary intakes.

4.4.1 Food items

Alcohol consumption

To my knowledge the associations between higher reported wine intake and increased branched-chain amino acids (BCAA; valine, leucine and isoleucine and their metabolites, 3-methyl-2-oxobutyrate and 4-methyl-2-oxopentanoate) and medium-chain fatty acids (caprate and caprylate) are novel findings. BCAA, 3-methyl-2-oxobutyrate and 4-methyl-2-oxopentanoate have been found to be higher in twins with type 2 diabetes or impaired fasting glucose (3-methyl-2-oxobutyrate being the strongest predictor) previously (Menni et al., 2013a) and associated with a higher BMI (Moore et al., 2014). Under impaired mitochondrial oxidation of lipids and glucose, levels of these BCAA catabolites increase, suggesting mitochondrial dysfunction. Insulin resistance has found to occur under binge drinking (Lindtner et al., 2013) though long-term effects are not known, these associations may highlight a novel pathway mediating this effect.

I also confirmed a number of metabolite associations with reported alcohol intake from a previous metabolomics study (Zheng et al., 2014) including higher intakes of wine with increased alpha-hydroxyisovalerate, scyllo-inositol, a inositol metabolite, and steroids originating from dehydroepiandrosterone (DHEA; 5-alpha-androstan-3beta,17beta-diol disulfate, 4-androsten-3beta,17beta-diol disulfate 1, 5-alpha-androstan-3beta,17beta-diol disulfate and epiandrosterone sulfate) (Zheng et al., 2014). Interestingly, a variant in the *HAO2* gene (rs12141041) encoding long-chain L-2-hydroxy acid oxidase 2 associated to alpha-hydroxyisovalerate. The *HAO2* gene has been found to be related to blood pressure regulation in animal models (Barawkar et al., 2012). A variant in the *SLC5A11* gene (rs4787294) is associated with scyllo-inositol. The *SLC5A11* gene encodes a myo- and scyllo-inositol

transporting sodium-dependent glucose transporter. The *SLC5A11* gene has been implicated in systemic lupus erythematosus (SLE) risk. Two SNPs the *SULT2A1* (rs2547231 and rs296396) gene which is involved in the sulfation of a number of steroids and bile acids, were associated to 4-androsten-3beta,17beta-diol disulfate 1 and 5-alpha-androstan-3beta,17beta-diol disulfate levels. Ethanol feeding in rats has recently shown significantly increased expression of *SULT2A1* in the intestines and liver (Maiti and Chen, 2015), suggesting this gene may be directly involved in modulating this association.

Seafood consumption

I identified novel associations between increased reported fish and seafood intakes with reduced pro-inflammatory lysolipids (1-arachidonoylglycerophosphoethanolamine, 1-eicosatrienoylglycerophosphocholine, 1-linoleoylglycerophosphoethanolamine, 1-oleoylglycerophosphoethanolamine), originating from essential fatty acid (EFA) metabolism. Lysolipids are involved in the formation of the cellular lipid bilayer, though contribute to inflammatory processes forming free lysophosphatidylcholines when cleaved by lipoprotein-associated phospholipase A2. Lysophosphatidylcholines potentially aggravate inflammatory artherosclerotic plaques (Goncalves et al., 2012). I replicated previous associations between increased reported intakes of oily fish and other seafood with higher 3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF), a furan fatty acid, and docosahexaenoate (DHA), an EFA (Guertin et al., 2014b, Zheng et al., 2014). Variants in *FADS1* (rs174538, rs174556, rs968567 and rs174535), a gene encoding a delta-5 desaturase enzyme, were associated with EFA lysolipids and DHA (Yang et al., 2015). Recently, a meta-analysis confirmed two of these variants in *FADS1* (rs174538 and rs174548) modulated gene-dietary-derived EFA associations (Smith et al., 2015).

Meat consumption

I found reported meat consumption to be associated with amino acids primarily, including creatine, trans-4-hydroxyproline and pyroglutamine. In particular, I found increased red meat intake to be associated with higher trans-4-hydroxyproline, a component of collagen previously shown to increase in blood following gelatin feeding (Ohara et al., 2007). Another novel association I identified was between increased reported red meat and poultry intakes and higher creatine levels. Red meat has characteristically high amounts of creatine and lower blood levels

have been found in vegetarians (Delanghe et al., 1989). Lower blood creatine has been associated with increased insulin sensitivity (Gall et al., 2010) and in liver steatotic versus NASH patients (Kalhan et al., 2011), higher levels have been found in individuals with dilated cardiomyopathy (Alexander et al., 2011). A variant in carbamoyl-phosphate synthase 1 (*CPS1*; rs715) was found to associate with creatine. CPS1 is the first enzyme of the urea cycle, producing ammonia from urea. Expression of *CPS1* is a candidate biomarker of NAFLD (Rodriguez-Suarez et al., 2012). Finally, higher intakes of seafood and meat correlated with lower levels of pyroglutamine, though little is known about this metabolite it is a cyclic derivative of glutamine, lower levels have previously been associated with chicken intake (Guertin et al., 2014b). A variant in *SLC6A13* (rs11613331), a gene that which encodes GAT2 a gamma-aminobutyric acid and betaine transporter, was associated with pyroglutamine. Other SNPs in *SLC6A13* have been related to renal function (Franceschini et al., 2014, Liu et al., 2011).

Dairy consumption

I identified novel associations between higher reported intakes of creams (double and clotted cream) with increased levels of lysoPhosphatidylcholine acyl C17:0 and C28:1 and hydroxysphingomyeline C14:1. A previous study on the EPIC-Potsdam cohort showed a diet pattern high in butter and high-fat dairy products and low margarine intake to be positively associated with lysoPhosphatidylcholine acyl C17:0 (Floegel et al., 2013b). Heptadecanoid acid (C17:0) has shown to be elevated following a dairy intervention trial, the lipid is formed by ruminal bacteria and therefore thought to be milk fat specific (Abdullah et al., 2015).

Reported butter consumption was uniquely associated with elevated levels of six lipids, including associations with increased butter intake and nonanoylcarnitine, an ester of carnitine with pelargonic acid (C9), 10-nonadecenoate (19:1n9), a monounsaturate of nonadecenoate (19:0), and myristate (14:0), a saturated fatty acid found most animal and vegetable fats. A variant in *ACADL* (rs3738934), a gene with important roles in lipid oxidation was associated with blood nonanoylcarnitine. Expression of *ACADL* has shown to be lower in mouse liver and adipocytes following high fat feeding, this effect also reduced overexpression of *Il-15* (Sun and Liu, 2015). I also confirmed associations with butter intake and pentadecanoate (15:0), 10-undecenoate (11:1n1) and 15-methylpalmitate from previous metabolomics studies (Guertin et al., 2014b, Zheng et al., 2014).

Low fat milk intake was strongly associated with a recently identified metabolite trimethyl-N-aminovalerate (X-21365). I replicated this association between general milk intake and trimethyl-N-animovalerate in the KORA population. The origin of this novel association is complicated to decipher as literature is limited on trimethyl-N-aminovalerate. Trimethyl-N-aminovalerate is potentially a methylated product of 5-aminovalerate, a product of lysine or proline degradation by gut bacteria *Escherichia coli* (Li et al., 2016) and *Pseudomonas putida* (Revelles et al., 2005). Moreover, both molecules are structurally similar to carnitine. Further research will be needed to unravel the mechanism of this association.

Grain-rich product intake

I found higher reported intakes of high fibre breakfast cereals to be associated uniquely with increased levels of pyridoxate, a vitamin B6 metabolite. Pyridoxate has many essential roles including being a coenzyme in the synthesis of amino acids, sphingolipids, neurotransmitters (serotonin, norepinephrine) and aminolevulinic acid. Higher pyridoxate in bloods has been associated with increased reported consumption of vitamins/supplements and other fruits (including plums, apricots, peaches, prunes, raisins, grapes and pineapple), as well as increased scores on the Healthy Eating Index (Guertin et al., 2014b). As it is common practice for breakfast cereals to be fortified with B vitamins this may have contributed to higher levels. I also found increased porridge intake to be associated with higher levels of 2-aminophenol sulphate. Previously, higher urinary output of 2-aminophenol sulphate was found in subjects fed whole grain rye bread versus refined wheat bread in a cross-over intervention study (Bondia-Pons et al., 2013). 2-aminophenol sulfate is a benzoxazinoid metabolite. Whole grains contain higher levels of benzoxazinoids which are readily absorbed from these products (Jensen et al., 2015). Higher reported intakes of refined grain products were associated with increased octenoylcarnitine. Octenoylcarnitine is an acylcarnitine formed from mitochondrial beta-oxidation. A previous metabolomics study undertaken on 33 coeliac disease patients showed those following a long-term gluten free diet had reduced octenoylcarnitine (Bene et al., 2005).

Fruit consumption

Higher reported intakes of apples and pears were associated uniquely with a sugar alcohol, threitol, and two amino acids formed from gut bacterial metabolism of phenolic compounds: indolepropionate (also associated with bananas) (Smith and Macfarlane, 1996, Karbownik et al.,

2001) and 3-phenylpropionate (Smith and Macfarlane, 1996). Apple proanthocyanidins have shown to be directly catabolized to 3-phenylpropionate when incubated with human gut bacteria (Ou et al., 2014). A previous study showed higher reported intakes of eggs and red meat in a US population to associated with lower indolepropionate in blood (Guertin et al., 2014b). Their finding may suggest these subjects had lower intakes of fruit-derived proanthocyanidins or higher consumption of animal proteins, which have been shown to rapidly alter the gut microbiome composition (David et al., 2014), and hence may indirectly influence bacterial metabolism of proanthocyanidins through altering the intestinal environment. Previously, increased indolepropionate has been associated with insulin sensitivity (Gall et al., 2010), and lower levels associated with reduced muscle mass in elderly subjects (Lustgarten et al., 2014). For 3-phenylpropionate (*ACSM5*, rs11647589) and indolepropionate (*ACSM2A,* rs1394678) SNPs within medium-chain acyl-CoA synthetase (MACS) genes were associated with levels. The ligation of medium-chain fatty acids with CoA is catalysed by MACS genes, though MACS also conjugate glycine with benzoic acid derivatices (Kasuya et al., 1996). Similar to 3-phenylpropionate and indolepropionate, benzoic acid is produced by gut microbial degradation of apple and cranberry phenolic compounds (Ou et al., 2014), the genotypic association with 3-phenylpropionate and indolepropionate may be due to their correlation to products of this process. An association with metabolic syndrome phenotypes and a variant in the *ACSM2* gene have been found previously; though this this association may be due mainly to the role of *ACSM2* in lipid metabolism (Lindner et al., 2006). I confirmed a previous association between reported intakes of fruit juice with stachydrine (also proline betaine) (Guertin et al., 2014b, Zheng et al., 2014), a component found in citrus fruits.

Vegetable consumption

Reported mushroom intake was associated with ergothioneine, a novel finding. Ergothioneine is a thiol compound found in higher quantities in specialty mushrooms such as oyster and king bolete (Ey et al., 2007). Ergothioneine has also shown to prevent lipid peroxidation *in vivo* (Deiana et al., 2004) and recent data suggest that ergothioneine may prevent vascular dysfunction (Li et al., 2014). I found higher reported intakes of green/leafy vegetables and avocado to associated with higher levels of metabolites that are strongly associated with seafood consumption including CMPF (3-carboxy-4-methyl-5-propyl-2-furanpropanoate) and 1-docosahexaenoylglycerophosphocholine, this issue may have arisen due to highly correlated

intakes, a problem encountered previously (Guertin et al., 2014a, Zheng et al., 2014, Ling et al., 2014).

Tea and coffee

I identified higher reported intakes of herbal tea were associated with lower levels of long-chain phosphatidylcholine acyl-alkyls derived from hepatic lipid metabolism (Wittenbecher et al., 2015). A dietary pattern characterised by a high intake of red meat and fish and lower intakes of whole grain bread and tea was previously shown to correlate with reduced levels of phosphatidylcholine diacyl (including phosphatidylcholine diacyl C42:4) in the EPIC Potsdam cohort (Floegel et al., 2013b). Also in the EPIC Potsdam cohort, levels of these phosphatidylcholines have associated with diabetes risk (Floegel et al., 2013a), though little is known currently about the origin of these associations.

I also confirmed associations between increased reported coffee intake and higher levels of metabolites derived from caffeine and coffee (Guertin et al., 2014a, Zheng et al., 2014). Interestingly, I also found a number of metabolites derived from coffee were associated inversely with reported black tea intake. This pattern is suggestive that individuals who reported having a higher tea intake had habitually lower coffee consumption. I identified higher levels of O-methyl catechol sulfate to be associated with higher reported coffee intake, which was a novel finding. One of the strongest metabolites associated to coffee intake was 1-methylxanthine. 1-methylxanthine is derived from caffeine metabolism and was previously associated to a variant in *NAT2* (rs4921914) which encodes a liver enzyme that acetylates caffeine metabolites (Butler et al., 1992). A SNP in *NAT2* has found to modulate an association with black tea consumption and risk of SLE (Kiyohara et al., 2014).

Sweet & savoury discretionary food intakes

Multiple associations were identified between foods where reduced consumption is promoted ('discretionary'; such as sweet baked products, sweets and jams, crisps, fried potatoes and fish and savoury pies) that seem to not be biologically sound and are also novel. This could suggest that these associations are showing habitual reduced intakes of other foods (i.e. vegetables, fruit, fish and wine) with increased reported intakes of discretionary foods. For instance, increased reported consumption of sweet baked products (including cookies, cakes, pies and pastries) were associated with lower levels of essential fatty acids derived from fish, (DHA and

EPA) and scyllo-inositol (strongly correlated with wine intake). Also, I found higher intakes of fried fish and savoury pies to be associated to lower levels of 3-phenylpropionate derived from gut microbial catabolism of fruit proanthocyanidins (Ou et al., 2014). Reporting of these foods may have been influenced by self-reporting biases (Westerterp and Goris, 2002), a limitation of this dataset.

Other notable associations

I identified a novel association with increased soymilk consumption and higher levels of 4-ethylphenylsulfate, despite intakes being low in the sample; a previous study of US subjects found higher 4-ethylphenylsulfate to be correlated with increased tofu consumption (Guertin et al., 2014b). 4-ethylphenylsulfate is formed under gut bacteria metabolism. Elevated levels of 4-ethylphenylsulfate have been found to correlate with anxiety-like traits in rats. Increased levels of 4-ethylphenylsulfate in blood are thought to be due to higher gut permeability (Hsiao et al., 2013). This association may be derived from the high saponin content of soybeans, which have shown to increase intestinal cell permeability in vitro (Johnson et al., 1986) though also in Atlantic salmon (Knudsen et al., 2008).

I found a novel association between higher reported chocolate intake and increased 7-methylxanthine, a methylated purine, and I also confirmed an association with theobromine, a bitter alkaloid and recognised marker of cocoa intake (Rodriguez et al., 2015). 7-methylxanthine is derived from methylxanthine metabolism, methylxanthines include caffeine, theophylline and theobromine (Suzuki and Takahashi, 1975).

Notable unknown metabolite associations

The chemical identity of 31 metabolites associated to food intakes are currently unknown, though these may become important food intake biomarkers in the future. In particular I found an association between increased reported fried food intake (fried fish and savoury snacks [including potato crisps]) and higher levels of the metabolite X-11372. I also identified another metabolite associated positively with red and processed meat consumption, metabolite X-11381. Metabolite X-11381 was previously associated with a SNP in a gene encoding a carnitine efflux transporter *SLC16A9* (rs12356193) (Suhre et al., 2011). Variants in *SLC16A9* have previously been associated with serum uric acid levels (Lee and Song, 2012) and risk of gout in renal overload (Nakayama et al., 2013). Another unknown metabolite, X-09789, was

associated positively with porridge intake. A variant in the *SLC51A* gene (also known as *OST-alpha*; rs7642243) was found to be associated with X-09789. Interestingly, *SLC51A* encodes a component of the Ost-alpha/Ost-beta complex which functions in bile acid transport from intestinal enterocytes into portal blood (Ballatori et al., 2008). Oats are known to contain beta-glucan, a soluble fibre found to reduce blood cholesterol levels (Charlton et al., 2012) by sequestering intestinal bile acids and effectively reducing the reabsorption of bile acids (Wolever et al., 2010). I also found the unknown metabolite X-11315 to be associated with reported consumption 13 foods (top association: apple and pears; negative associations with discretionary foods). Metabolite X-11315 was previously associated to a SNP in *SLC6A20* (rs4327428), which encodes a proline transporter, this suggests metabolite X-11315 is structurally like proline. Previous studies have identified polymorphisms in *SLC6A20* to be associated with risk of Type 2 diabetes in white-European and Chinese populations (Ling et al., 2014).

4.4.2 Dietary patterns discussion

Examining dietary patterns in conjunction with metabolomics, I identified a number of potential metabolite biomarkers of dietary intakes in the twin population and validated the results using the co-twin control method. The repetitive occurrence of a number of metabolites within opposing dietary patterns in this study (**Table 4-5**) further confirms the utility of metabolomics for the identification and exploration of diet and disease relationships. I also identified a number of metabolites unique to each diet pattern (**Table 4-6)**. Many of the associations with dietary patterns were attributable to their association with food intakes, therefore I will not discuss these associations in detail. However, there were 11 known metabolites and 10 unknown metabolites identified by measuring dietary patterns that I did not find to associate with food intakes independently. The Traditional English and Dieting patterns did not appear to have a substantial metabolomic footprint and did not capture unique associations with metabolites, this may be due to underreporting.

Fruit and vegetable pattern

Twins scoring highly on the F&V pattern presented with an increase in metabolites relating to antioxidant potential, including hippurate which accumulates following increased consumption of

foods containing high phenolic compounds such as green and black teas (Van Dorsten et al., 2006); threonate, a product of vitamin C metabolism; and the **γ**-glutamyl peptide, **γ**-glutamylvaline and proline which both suggest reduced glutathione and collagen breakdown on this diet. Higher scores on the F&V diet pattern were also associated with increased phosphatidylcholine diacyl 38:6 and N-acetylornithine. N-acetylorthinine is an endogenously synthesized amino acid involved in urea metabolism. A genetic variant in N-acetyltransferase 8 gene (*NAT8*) was found to contribute to N-acetylornithine levels which may play role in kidney function and CKD (Yu et al., 2014). Despite the MDS and F&V patterns characterised by similar foods the data-driven F&V score captured unique metabolite associations, suggesting that measuring a Mediterranean diet in a non-Mediterranean country such as the UK may be problematic, an issue which has been thoroughly discussed previously (Hoffman and Gerber, 2013).

Low meat dietary pattern

Higher scores of the low meat dietary pattern were associated with lower levels of phosphatidylcholine acyl-alkyl 38:6 and higher levels of betaine and C-glycosyltryptophan. Betaine is an amino acid derived from choline. Choline is a nutrient that has an important role in methylation processes and reduction of homocysteine levels and subsequent risk of coronary heart disease (Zeisel and da Costa, 2009), the origin of this association requires further investigation. The low meat diet pattern was positively associated with C-glycoyltryptophan. Circulating levels of C-glycosyltryptophan have been associated with aging and age-related diseases in the TwinsUK population (Menni et al., 2013b), negatively associated with muscle mass in less mobile elderly subjects (Lustgarten et al., 2014), and a CpG site in the promoter of the WDR85 gene, which is required for protein synthesis (Menni et al., 2013b). This could suggest that individuals consuming a dietary pattern low in protein but high in starches and carbohydrates tend to be older or this type of dietary pattern may contribute to physical aging and reduced muscle mass.

High alcohol dietary pattern

A unique association was identified between the high alcohol diet and 4-vinylphenol sulphate, a product of benzoate metabolism and compound resulting from nut roasting (Walradt, 1971), and airborne styrene exposure and cigarette smoking (Manini et al., 2003), that has been previously

associated with nut consumption (Guertin et al., 2014b) and several CpG loci (Petersen et al., 2014). The metabolic implications for this epigenetic association are unknown at this time, though it may suggest that individuals scoring highly on this pattern are more prone to cigarette smoking. Increased scores on the alcohol pattern were also associated with higher levels of gamma-tocopherol, Gamma-tocopherol is one form of vitamin E, a major dietary antioxidant. The relationship between alcohol intake and gamma-tocopherol is unclear, a previous study found higher intakes of vitamins and supplements, and scores on the US Healthy Eating Index were associated with lower serum γ-tocopherol (Guertin et al., 2014b).

Table 4-5. List of metabolites associated to multiple dietary patterns

Super-pathway	Sub-pathway	Metabolite	MDS	F&V	High alcohol	Traditional English	Low meat
Amino acid	Creatine metabolism	**Creatine**				↑	↓
	Phenylalanine & tyrosine metabolism	**3-Phenylpropionate (hydrocinnamate)**		↑		↓	
	Tryptophan metabolism	**Indolepropionate**	↑	↑		↓	
	Valine, leucine and isoleucine metabolism	**α-Hydroxyisovalerate**			↑		↓
Carbohydrate	Glycolysis, gluconeogenesis, pyruvate metabolism	**1,5-Anhydroglucitol (1,5-AG)**		↓			↑
		Glycerate	↑	↑			
Lipid	Essential fatty acid	**Docosahexaenoate (DHA; 22:6n3)**	↑	↑			↓
		Eicosapentaenoate (EPA; 20:5n3)	↑	↑			↓
	Fatty acid, dicarboxylate	**3-Carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF)**	↑	↑			↓
	Inositol metabolism	**Scyllo-inositol**	↑	↑	↑		↓
	Lysolipid	**1-Docosahexaenoylglycerophosphocholine**	↑	↑			
	Medium chain fatty acid	**10-Undecenoate (11:1n1)**			↑		↓
Xenobiotics	Benzoate metabolism	**Catechol sulfate**		↑		↓	

Table 4-6. List of metabolites associated to specific dietary phenotype

Dietary phenotype	Super-pathway	Sub-pathway	Metabolite	Direction of association
Fruit & vegetable	Amino acid	Lysine metabolism	**Pipecolate**	↑
	Carbohydrate	Nucleotide sugars, pentose metabolism	**Threitol**	↑
	Cofactors and vitamins	Ascorbate and aldarate metabolism	**Threonate**	↑
	Peptide	gamma-glutamyl	**γ-Glutamylvaline**	↓
	Xenobiotics	Benzoate metabolism	**Hippurate**	↑
High alcohol	Amino acid	Butanoate metabolism	**2-Aminobutyrate**	↑
	Lipid	Inositol metabolism	**Myo-inositol**	↑
		Medium chain fatty acid	**Caprate (10:0)**	↑
		Medium chain fatty acid	**10-Undecenoate (11:1n1)**	↑
	Xenobiotics	Food component/Plant	**Piperine**	↑
Traditional English	Amino acid	Urea cycle; arginine-, proline-, metabolism	**Stachydrine**	↓
		Urea cycle; arginine-, proline-, metabolism	***Trans*-4-hydroxyproline (Hydroxyproline)**	↑
Dieting	Xenobiotics	Benzoate metabolism	**4-Ethylphenylsulfate**	↑
		Food component/Plant	**Quinate**	↓
Low meat	Amino acid	Glutamate metabolism	**Pyroglutamine**	↑
		Glycine, serine and threonine metabolism	**Betaine**	↑
		Tryptophan metabolism	**C-Glycosyltryptophan**	↑
		Cysteine, methionine, SAM, taurine metabolism	**2-Hydroxybutyrate (AHB)**	↓

Strengths and Limitations

Although I replicated many associations from previous diet-metabolomics studies, there were multiple limitations to my study. The population I used was only female and as such my findings may not apply to men. Replicating my top associations in an independent population would have strengthened my findings, although I used the twin model for validation, allowing me to use controls matched for age, sex and the baseline genetic sequence. I did however, confirm associations from similar studies (Guertin et al., 2014b, Zheng et al., 2014) which suggests the data used in the study were of decent quality. Due to the cross-sectional nature of this study design, the study does not allow me to indicate cause and effect. As I used only FFQ data as my source for dietary information, the accuracy of the data can be called into question due to the consequences of self-reporting (Westerterp and Goris, 2002). Despite this issue, many of the associations I identified are biologically plausible. Moreover, food frequency data only provides categorical information and therefore the precise effect cannot be quantified, though this is an area where future dietary intervention studies can aid findings. I do recognize I may have experienced a small number of associations due to type 1 errors potentially as a result of correlated intake reporting, this issue may have been encountered for foods which lower intakes are generally discouraged, including sweet and savoury discretionary foods (sweets and jams, sweet baked products, fried fish and savoury pies). Using very stringent multiple-testing cut-offs I am hopeful many of these spurious associations were minimized, although this likely caused false negatives. Moreover, I did not have longitudinal data on metabolite levels, which in future studies could provide useful information on metabolite stability over time and may improve the strength of associations with food intakes. The true blood concentration of the metabolites is not measured by the Metabolon platform, therefore accurate quantification could not be determined for these metabolites. Furthermore, the current study only provides a snapshot of associations in one type of tissue, at only one time point and therefore provides limited insight into the origin and directionality of associations and virtually no insight into mechanisms. Having genotypic profiling completed on a large sample of TwinsUK with results also replicated in KORA, this allowed me to supplement my discussion, though future studies would need to untangle these potential genotype-metabolite-disease relationships and whether they are modulated by dietary factors. Strengths of this chapter include the large sample size and number of metabolites that were assessed and I was also able to use MZ discordant twins who are matched for age, sex and the baseline genetic sequence.

Conclusions

By using one of the largest and comprehensive datasets of its kind, I identified 178 self-reported food intake associations (72 novel) with blood metabolites. I also identified 108 associations with 6 dietary patterns, confirming dietary patterns are associated with a unique metabolomic signature not captured by assessing food intakes independently. Future studies should aim to undertake dietary interventions trials to confirm our findings, adequately determine mechanisms for associations and quantify the effect of food intake on metabolite levels. The findings of my study can be viewed online using the DietMetab search tool (http://www.twinsuk.ac.uk/dietmetab-data/).

Chapter 5 Creating metabolite scores based on reported food group intake and applying these to explore the impact of diet on the metabolic syndrome

In this chapter I analyse 20 food groups for their association with metabolites from the non-targeted (Metabolon) platform, I then use three methods to combine multiple metabolites associated with food group intakes and evaluate them in an independent testing dataset. Therefore the overall aim of this chapter is to assess combining metabolites as a means to strengthen the utility of food intake biomarkers. Finally, I use the top performing metabolite scores to evaluate their correlation with an example of a common disorder - metabolic syndrome risk.

5.1 Introduction

In the previous chapter I showed that reported food intakes are associated with a unique blood metabolomic fingerprint. In order to best categorise habitual food intake to improve disease risk prediction, the appropriate handling of identified biomarkers must first be evaluated.

Used alone single biomarkers may be less precise as they can be influenced by a multitude of factors and non-specific to a particular disease state. For example, c-reactive protein is synthesised by the liver under inflammatory states and is therefore used as a general marker for inflammatory processes occurring in the body and not specific to one condition. Indeed it is one marker of the metabolic syndrome (Olza et al., 2015) and CVD risk (Berezin et al., 2015). Moreover, complex chronic diseases are characterised by clusters of symptoms that cannot be quantified by measuring a single biomarker. Biomarker risk scores have been evaluated and shown to improve accuracy over traditional methods for a multitude of conditions, such as CVD (Berezin et al., 2015, Richter et al., 2013, Hughes et al., 2012) and metabolic syndrome (Olza et al., 2015).

Similarly, combining dietary information to create scores is currently performed, such as for creating dietary patterns *a priori* to measure adherence to dietary patterns associated with particular health outcomes or measuring diet quality, including the MDS (Trichopoulou et al., 2003) and the Healthy Eating Index (Guenther et al., 2013). Scoring is assigned using cut-offs

points previously associated to health outcomes or diet quality. Diet quality is indicated by scoring reported food intakes by how closely they adhere to established dietary guidelines and also the variety of healthy food choices within core food groups. Dietary adherence scores have been tested for their correlation with traditional biomarkers of nutrient intake (e.g. n-3 fatty acids and carotenoids) (Golley et al., 2015). A study from EPIC-Norfolk created a panel of traditional biomarkers of fruit and vegetable intakes by summing plasma vitamin C, beta-carotene and lutein and found the panel to be strongly inversely associated with incident type II diabetes (Cooper et al., 2015). Moreover, a multi-food intake biomarker approach applying summation and PCA methods was used to evaluate adherence to a healthy Nordic diet dietary intervention in patients with MetS (Marklund et al., 2014), both methods performed similarly. With emerging nutritional metabolomics studies measuring multiple biomarkers simultaneously these traditional biomarkers may soon be replaced by more precise food intake scores.

In one nutritional metabolomics study biomarkers of alcohol consumption in African Americans were identified and combined to create an alcohol risk score to evaluate the effects of alcohol consumption on levels of white blood cells and incident CVD (Zheng et al., 2014) . Specifically, the authors created three composite scores by summing quartile-ranked metabolites belonging to the same alcohol-related metabolic sub-pathways, including 2-hydroxybutyrate and related BCAAs, lysophosphatidylcholines, and γ-glutamyl dipeptides. However, combining this information may have been more useful as independently these pathways are affected by different health conditions. In another study, a metabolomics approach was used to identify a panel of metabolites related to sugar-sweetened beverage (SSB) consumption, where a panel of 4 metabolites (formate, citrulline, taurine, and isocitrate) were found to be discriminate consumers and non-consumers (Gibbons et al., 2015).

Combined biomarker scores representing long-term intakes of frequently consumed foods may greatly improve our understanding of the health impacts of diet. Therefore, my aims for this chapter were to:

i) Identify potential biomarkers of intakes of 20 commonly consumed food groups.

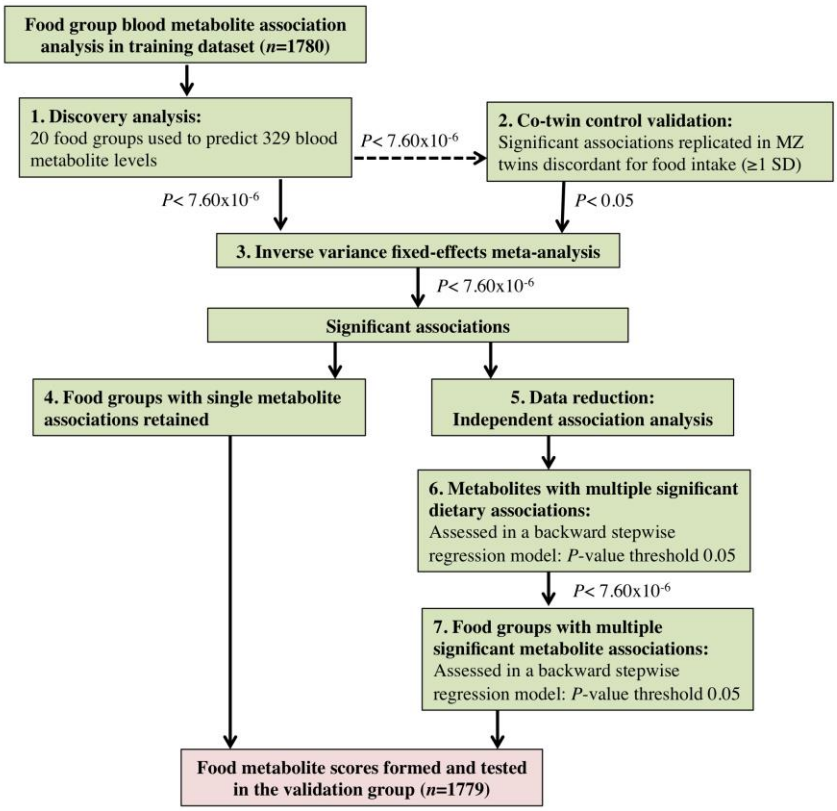
ii) Form food metabolite scores using multiple methods and test their utility.

iii) Use the scores to predict the risk of developing the metabolic syndrome.

5.2 Materials and methods

Figure 5-1 shows the pipeline of the chapter.

Figure 5-1. Study pipeline of the chapter



5.2.1 Food groups

I created and analysed twenty food groups of similar taste and nutrient content (shown in **Table 5-1**) derived from FFQ against blood metabolites**.**

Table 5-1. List of food items included in food groups

Food group	FFQ items
Vegetables	
	Broccoli, spring green, kale
	Brussel sprouts
	Cabbage
	Cauliflower
	Coleslaw
	Avocado
	Beetroot
	Marrow, courgettes
	Mushrooms
	Parsnips, turnips, swedes
	Sweetcorn
	Sweet peppers
	Watercress
	Carrots
	Tomatoes
	Garlic (clove)

Table 5-1. List of food items included in food groups

Food group	FFQ items
	Leeks
	Onions
	Green salad, lettuce, cucumber, celery
	Spinach
	Watercress
	Vegetable soups (bowl)
	Boiled, mashed, instant or one jacket potato
Fruit	
	Strawberries, raspberries, other berries, kiwi fruit (one fruit or handful)
	Smoothies (cup)
	Pure fruit juice (100%) e.g. orange, apple juice (cup)
	Grapefruit (half)
	Oranges, satsumas, mandarins (1 fruit)
	Apples (1 fruit)
	Bananas (1fruit)
	Dried fruit, e.g. raisins, prunes (heaped tablespoon)
	Grapes (handful)
	Melon (1 slice)
	Peaches, plums, apricots (1 fruit)
	Pears (1 fruit)
	Tinned fruit (handful)
Whole grains	
	High Fibre cereals e.g. Branflakes, All Bran, Fruit and Fibre
	Muesli
	Porridge, Readybreak, oats
	Brown rice
	Wholemeal & granary bread/rolls
	Wholemeal pasta
	Crispbread, e.g. Ryvita
Refined grains	
	Breakfast cereal e.g. Cornflakes, Rice Krispies
	Sugar topped cereals e.g. Frosties
	Naan, poppadoms, flour tortillas
	Brown bread/rolls
	White bread/rolls
	White or green pasta, e.g. spaghetti, macaroni, noodles
	White rice
Nuts and legumes	
	Beansprouts
	Pulses e.g. lentils, beans, peas
	Green beans, broad beans, runner beans
	Peas
	Baked beans
	Salted nuts e.g. peanuts, cashews (handful)
	Unsalted nuts, e.g. brazil, walnuts (handful)
	Seeds e.g. Sunflower, pumpkin (tablespoon)
	Peanut butter (teaspoon)
	Meat substitutes e.g. tofu, soyameat, textured vegetable protein, vegeburger
Seafood	
	Oily fish, fresh or canned, e.g. tuna, mackerel, kippers, salmon, sardines, herring
	Fish roe, taramasalata
	Shellfish, e.g. crab, prawns, mussels
	Other white fish, fresh or frozen, e.g. cod, plaice, sole, haddock, halibut
White meat	
	Chicken or other poultry e.g. turkey
Red meat and eggs	
	Beef: roast, steak, mince, stew or casserole
	Lamb: roast, chops or stew
	Pork: roast, chops or stew
	Beefburgers
	Bacon or gammon
	Corned Beef, Spam, luncheon meats
	Ham, cured meats & chorizo

Table 5-1. List of food items included in food groups

Food group	FFQ items
	Liver, liver pate, liver sausage
	Sausages
	Eggs as boiled, fried, scrambled, etc. (one)
Fermented dairy	
	Low fat cheese e.g. reduced fat cheddar (matchbox size)
	Cheese, e.g. cheddar, brie, edam (matchbox size)
	Cottage cheese, low fat soft cheese (2 tablespoons)
	Full fat or Greek yoghurt (small pot)
	Low fat yoghurt, fromage frais (small pot)
Fried foods	
	Fish fingers, fish cakes & breaded fish
	Fried fish in batter, as in fish and chips
	Chips, retail, fried in vegetable oil
	Potato salad
	Old potatoes, roast in blended oil
	Savoury pies, e.g. meat pie, pork pie, pasties, steak & kidney pie, sausage rolls
	Cream crackers, savoury biscuits
	Crisps or other packet snacks, e.g. Wotsits (one packet)
	Pizza (one slice)
	Quiche (slice)
Sweets and sweet baked products	
	Reduced fat biscuits e.g. Go Ahead, Highlights (one small packet or one small bar/biscuit)
	Sweet biscuits, chocolate, e.g. digestive (one)
	Sweet biscuits, plain, e.g. Nice, ginger (one)
	Buns, pastries e.g. scones, flapjacks, croissants, doughnuts, home baked
	Cakes e.g. fruit, sponge, home baked
	Cakes e.g. fruit, sponge, ready made
	Fruit pies, tarts, crumbles, home baked
	Fruit pies, tarts, crumbles, ready made
	Milk puddings e.g. rice, custard, trifle
	Sponge puddings, home baked
	Sponge puddings, ready made
	Dairy desserts (small pot) e.g. chocolate mousse, cream caramels
	Ice cream, choc ices
	Jam, marmalade, honey (teaspoon)
	Sugar added to tea, coffee, cereal (teaspoon)
	Sweets, toffees, mints (small packet)
Chocolate	
	Dark chocolates, single or squares (one)
	White or milk chocolates, single or squares (one)
	Low fat hot chocolate (cup)
	Cocoa, hot chocolate (cup)
	Chocolate snack bars e.g. Mars, Crunchie (one)
Butter and cream	
	Reduced fat butter (teaspoon)
	Butter (teaspoon)
	Double or clotted cream (tablespoon)
	Single or sour cream (tablespoon)
Spreads and dressings	
	Low fat spread, e.g. Outline, Gold (teaspoon)
	Very low fat spread (teaspoon) e.g. Diet Flora
	Cholesterol lowering fat spreads e.g. Benecol (teaspoon)
	Olive oil spread (teaspoon)
	Block margarine, e.g. Stork, Krona (teaspoon)
	Other soft margarine, dairy spreads, e.g. Blue Band, Clover (teaspoon)
	Polyunsaturated margarine, e.g. Flora, sunflower (teaspoon)
	French dressing (tablespoon)
	Full fat salad cream, mayonnaise (tablespoon)
	Other salad dressing (tablespoon)
	Low calorie, low fat salad cream (tablespoon)
Milk	
	Channel Islands milk
	Full cream milk

Table 5-1. List of food items included in food groups

Food group	FFQ items
	Dried milk
	Semi-skimmed milk
	Skimmed milk
Soya and other milk	
	Goats' milk
	Rice milk
	Soya milk
Soda	
	Fizzy soft drinks, e.g. Coca Cola, lemonade (cup)
	Low calorie or diet fizzy soft drinks (cup)
Tea	
	Tea (cup)
	Green tea (cup)
Coffee	
	Coffee, instant or ground (cup)
	Coffee, decaffeinated (cup)
Alcohol	
	Beer, lager or cider (half pint)
	Port, sherry, vermouth, liqueurs (pub measure)
	Spirits, e.g. gin, brandy, whisky, vodka (pub measure)
	Red wine (small glass)
	White wine (small glass)

5.2.2 Metabolomics dataset

Fasted blood metabolites derived from the Metabolon dataset were used (**Section 3.1.6.1**). I included in the analysis 329 metabolites that were present in both serum and plasma and had no missing values in 75% of the sample.

5.2.3 Training and validation datasets

Female twins with Metabolon data available within and including 5 years of FFQ completion were included in the total sample (*n*=3559). A subset of 1156 twins had food preference data available, I therefore randomly assigned 1779 twins (keeping co-twins together) of this subset (50% of the whole sample) to the validation group, while the remaining 50% of the sample (*n*=1780) were assigned to the training group. **Table 5-2** shows the characteristics of the training and validation datasets.

Table 5-2. Characteristics of the training and validation groups

	Training group				Validation group	
	Discovery[1]		MZ discordant			
Phenotype	***n***	**Mean (SD)**	***n***	**Mean (SD)**	***n***	**Mean (SD)**
Age (years)	1780	56.9 (14.9)			1779	54.4 (11.6)
BMI (kg/m2)	1766	26.1 (4.8)			1771	26.1 (4.9)
Food groups (servings/week)[2]						
Vegetables	1652	32.1 (16.2)	128	40.4 (24.1)	1779	34.2 (15.5)
Fruit	1588	20.3 (12.7)	192	26.2 (16.9)	1779	21.8 (12.5)
Whole grains	1582	9.7 (8.6)	198	14.4 (11.2)	1779	10.3 (8.2)

Table 5-2. Characteristics of the training and validation groups

	Training group				Validation group	
	Discovery[1]		MZ discordant			
Phenotype	***n***	**Mean (SD)**	***n***	**Mean (SD)**	***n***	**Mean (SD)**
Refined grains	1588	10 (8.3)	192	13.9 (11.2)	1779	9.6 (8.2)
Nuts and legumes	1598	6.5 (4.3)	182	9.9 (6.1)	1779	7.5 (5.3)
Seafood	1590	2.1 (1.8)	190	3.5 (2.7)	1779	2.4 (1.9)
White meat	1492	1.8 (1.3)	288	2.2 (1.4)	1779	2.0 (1.3)
Red meat	1626	6.5 (3.7)	154	8.5 (5.1)	1779	6.7 (3.8)
Fermented dairy	1618	5.5 (4.3)	162	8.2 (7.2)	1779	6.2 (4.8)
Fried foods	1604	4.9 (3.6)	176	7.3 (5.9)	1779	4.5 (3.3)
Sweets and sweet baked products	1618	17.4 (14.5)	162	27.0 (22.6)	1779	15.6 (14)
Chocolate	1662	3.3 (5.0)	118	7.8 (9.3)	1779	4.0 (6.0)
Butter and cream	1616	3.3 (5.7)	164	8.6 (8.8)	1779	4.1 (6.4)
Spreads and dressings	1610	8.8 (8.5)	170	14.9 (13.2)	1779	8.6 (9.1)
Milk	1592	3.7 (2.2)	188	4.4 (3.1)	1779	3.4 (2.2)
Soya and other milk	1752	0.1 (0.6)	28	2.0 (2.1)	1779	0.2 (0.8)
Soda	1672	2.3 (5.2)	108	7.8 (8.3)	1779	2.3 (5.5)
Tea	1584	20.2 (13.8)	196	19.7 (13.8)	1779	19.0 (13.8)
Coffee	1574	8.9 (10.8)	206	15.9 (12.6)	1779	8.9 (10.7)
Alcohol	1682	4.9 (7.1)	98	12.0 (11.7)	1779	6.3 (7.8)

[1]Excluding twins discordant for food group intake.
[2]Variables are residual adjusted for energy intake.

5.2.4 Classification of the metabolic syndrome

During clinical visits undertaken on the twins estimates of waist circumference were determined by analysing DXA scans, fasted blood samples were taken and analysed to determine levels of total triglycerides (TG), high density lipoprotein (HDL)-cholesterol and glucose, and blood pressure (BP) was determined, consult **Section 3.5.1** for details.

I determined MetS status using the criteria outlined by the International Diabetes Federation and the American Heart Association/National Heart, Lung, and Blood Institute (Alberti et al., 2009). Twins who had at least three of any of the following factors were considered to have the metabolic syndrome:

1. Increased waist circumference (WC): men ≥94 cm and women ≥80 cm, or BMI > 30 kg/m^2.
2. Increased triglycerides (TG): >1·7 mmol/L, or treatment for abnormality
3. Lower HDL-cholesterol: men <1·03 mmol/L in men and women <1·29 mmol/L, or treatment for abnormality
4. High blood pressure: ≥130/85 mm Hg, or treatment for hypertension
5. Raised fasting plasma glucose (FBG): FBG ≥ 5·6 mmol/L. Most individuals with type 2 diabetes will have metabolic syndrome based on these criteria

I included 414 twins in the analysis who had metabolomics profiling and diet information and did not have MetS at this time. These twins also attended a clinical visit 5 years or more after metabolomics profiling where MetS status was confirmed.

5.2.5 Statistical analysis

Statistical analysis was carried out using Stata version 12.

5.2.5.1 Discovery analysis in the training dataset

For each metabolite, random intercept linear regression analysis was undertaken in the first sample (discovery sample) excluding MZ twin pairs discordant (MZ twins with measures one SD apart in food group intake) for each food group. Age, metabolite batch, BMI and family relatedness were included as covariates:

$$Y_i = \beta_0 + \beta_i X_{ij} + \gamma_i age_{ij} + \delta_i BMI_{ij} + \zeta_j + \varepsilon_{ij}$$

where Y_i is the metabolite and X_{ij} the food group intake of twin j from pair i , ζ_j, is the family-specific error component that captures the unobserved heterogeneity or family characteristics.

I adjusted for multiple testing using Bonferroni correction thus giving a significant threshold of 7.60×10^{-6} (0.05/(20 food groups x 329 eligible metabolites)). For each significant metabolite-food group association from the discovery sample, I repeated the same linear regression analysis on the MZ discordant twin pair sample. The MZ discordant twin pairs were used to replicate the significant findings from the discovery group, associations in the same direction as the discovery group were considered replicated. Finally, we combined the results of both analyses using an inverse variance fixed effects meta-analysis that are the reported results. The beta coefficients (β) presented in the results of each linear regression analysis represent the amount of a food group consumed in servings per week that corresponds to a 1 SD change in the metabolite level.

5.2.5.2 Data reduction

Following identification of significant metabolites for each food I reduced the number of associations.

Firstly, a number of metabolites were associated with multiple food groups, which may have been a result of correlations in intake reporting. I therefore conducted a backward stepwise linear regression using the 5% level of significance as the cut-off threshold for each metabolite (adjusted for covariates) using all food groups associated as predictors of the

metabolite accordingly. I considered associations to be significant if they passed the Bonferroni cut-off for multiple testing used for the discovery analysis ($P<7.60x10^{-6}$).

Secondly, a number of the food groups were associated with multiple potentially correlated metabolites. Although the hypothesis is that metabolite levels change with food intake, to determine which metabolites are correlated independently of one another with food intake I undertook a backwards stepwise linear regression (cut-off threshold: $P<0.05$), with food group intake as the dependent variable. I performed this on all the metabolite residuals identified following the discovery and first data reduction steps, to identify a potential panel of independent metabolites to use as food intake markers.

5.2.5.3 Food-metabolite group generation

Significant food group-metabolite associations were then converged into scores based on three different methods: 1) Similar to the method used by Zheng and colleagues (Zheng et al., 2014), I quartile ranked blood levels of significant metabolites and assigned the quartiles a score of 0 to 3 according to the direction of the association (i.e. positive association: Q1=0, Q2=1, Q3=2, Q4=3; negative association: Q1=3, Q2=2, Q3=1, Q4=0) and summed these values; 2) Like the method used by Marklund and colleagues (Marklund et al., 2014), I created a continuous summed score by summing the relative levels of each significant metabolite. Negatively-associated metabolites were multiplied by -1; 3) I created a weighted score by multiplying the relative levels by the beta-coefficients for each metabolite as a predictor of food intakes (standardized to have mean 0 and SD 1), adjusting for all other associated metabolites and summed the final scores. This method has been used by Dash and colleagues (Dash et al., 2013) to examine the indicators of oxidative balance and incident, sporadic colorectal adenomas.

5.2.5.4 ROC analysis by food intake

To evaluate the utility of the food metabolite scores as potential markers of food intakes I conducted a binary classification test in the validation group. I classified twins in lower tertile of each food group as a negative outcome (0), and the top tertile of food group intake I considered a positive outcome (1). The ability of the metabolite scores to correctly classify twins consuming high intakes (sensitivity; true positive rate) and correctly classify twins consuming low intakes (specificity; true negative rate) of the model was predicted and the receiver operating characteristic curve (ROC) generated by plotting the true positive rate against the false positive

rate at multiple threshold settings. I tested whether any scoring method had a consistent advantage over another method by testing the difference in the area under the ROC curve (AUC). The AUC may be interpreted as follows: AUC>0.9 has high accuracy, AUC>0.7–0.9 has moderate accuracy, AUC>0.5–0.7 has low accuracy, and AUC=0.5 is a chance result (Fischer et al., 2003).

Confirmation of Associations by Food Preference

I used an alternative form of getting diet information from subjects by asking them their food preferences. I adapted the food preference questionnaire developed in the US by Dr Valerie Duffy (Duffy et al., 2007) for use in the UK twin cohort (details in **Appendix D Document 1**). To do this, in collaboration with other scientists, items were changed to suit the UK diet and I transformed the questionnaire into an online format, this work has been published (Pallister et al., 2015). Modifications to the questionnaire involved changing the terms for some items (e.g. 'French fries' to 'chips', 'breakfast sausage' to 'sausage', and 'unflavoured oatmeal' to 'porridge') and some items were removed as they are infrequently consumed in the UK (e.g. 'bologna', 'canned noodle soup', 'chocolate milk', and 'blackened hot dog'). I compared their preferences for foods (ranked on a scale from -100 to +100) against those food groups as a way to further validate the metabolite scores. **Table 5-3** shows the foods which ratings were averaged to form the group. I ran a linear regression with the metabolite score as the response variable (adjusted for batch effects, BMI and age) and twin food group liking as the predictor variable adjusting for family relatedness. Those associations passing the cut-off for multiple testing (Bonferroni $P<3.33x10^{-3}$) were considered validated.

Table 5-3. List of food preference items by food group

Food group	Preference item
Vegetables	
	White potato
	Broccoli
	Fresh tomatoes
	Fresh coriander
	Garlic
	Sautéed mushrooms
	Asparagus
	Raw carrots
	Beetroot
	Spinach/ greens
	Aubergine
	Raw onion
Fruit	
	Grapefruit
	Banana
	Lemon

Table 5-3. List of food preference items by food group

Food group	Preference item
	Pear
	Melon
	Cherries
	Strawberries
	Pineapple
	Orange juice
Whole grains	
	Wholemeal bread
	Porridge
	High fibre bar
Refined grains	
	Bagel/ rolls
	Pasta/ noodles
	Cornflakes
	White rice
Nuts and legumes	
	Unsalted nuts
	Faux meat products
	Lentils/ beans
Seafood	
	Tuna or salmon
	Prawns and shellfish
White meat	
	Baked chicken
Red meat	
	Crispy bacon
	Sausage
	Pork chops
	Ham
	Chargrilled meat
	Eggs
	Beef steak
Fermented dairy products	
	Plain yoghurt
	Blue cheese
	Cheddar cheese
Fried and fast foods	
	Tortilla chips or crisps
	Chips
	Fried chicken
	Fried fish
	Pizza
Sweets and sweet baked products	
	Ice cream
	Cake icing
	Jam and jelly
	Biscuits, cakes or pastries
	Cheesecake
Chocolate	
	Dark chocolate
Butter	
	Butter or margarine
Spreads and dressings	
	Butter or margarine
	Mayonnaise
	Salad dressing
	Extra virgin olive oil
Milk	
	Skimmed milk
	Whole milk
Soy milk	
	Soy milk

Table 5-3. List of food preference items by food group

Food group	Preference item
Soda	
	Carbonated drinks & sweet drinks
	Diet carbonated soft drinks
Tea	
	Hot tea
Coffee	
	Black coffee
	Coffee or tea with sugar
	Sweet coffee drinks & whipped cream
Alcohol	
	Beer
	Red wine
	White wine
	Vokda, gin or scotch

5.2.5.5 Usage of food metabolite scores to predict the metabolic syndrome

To determine whether food metabolite scores were associated with risk of developing MetS, I used the food metabolite scores that performed the best (weighted scores if multiple metabolites and continuous variable for foods associated to single metabolites). I undertook a logistic regression in 414 twins using each food metabolite score (residual adjusted for age, BMI and batch at time of metabolomics) to predict future MetS status (0, no MetS [*n*=389]; 1, MetS [*n*=25]) adjusted for age, BMI and family relatedness. Statistical significance was defined as $P<0.05$. I performed the same logistic regression on reported food intake for metabolite scores significantly predictive of future MetS. I finally performed a logistic regression using each significant metabolite score as a predictor of each of the 5 clinical criteria for MetS.

5.3 Results

In the following paragraphs I will first present those metabolite associations used to create the food metabolite scores from the training dataset, followed by the validation results.

5.3.1 Overall associations with food group intakes

A summary of the results for the discovery and data reduction analysis can be found in **Table 5-4**. Following the meta-analysis of the discovery and discordant twin groups, 112 significant associations were identified in the training dataset (**Figure 5-2**; **Appendix D Table 1**). Significant associations were identified for all food groups except chocolate and refined grain intake. Thirty-nine metabolites were associated only with one food group (unique associations; **Table 5-5**), whereas 26 metabolites were associated with multiple foods (73 associations overall) potentially as a result of correlated intakes. Following a multivariate backward stepwise linear regression using each associated food group as a predictor of the latter 26 metabolites (**Appendix D Table 2**), 35 associations (including 24 metabolites) remained significant after Bonferroni correction ($P<7.60 \times 10^{-6}$).

Table 5-4. Results summary of the discovery analysis, the 1st and 2nd data reductions, and the final number of metabolites contributing to each food-metabolite score with R^2

Food group	Total metabolites associated in discovery analysis (1)	No. metabolites removed after 1st data reduction (2)	No. metabolites removed after 2nd data reduction (3)	Final no. metabolites	R^2
Vegetables	3	2	0	1	0.0094
Fruit	10	2	4	4	0.1195
Whole grains	7	4	0	3	0.0508
Nuts and legumes	2	1	0	1	0.0384
Seafood	18	3	8	7	0.1535
White meat	1	0	0	1	0.0291
Red and processed meat, and eggs	4	1	0	3	0.0403
Fermented dairy	1	1	0	0	-
Fried foods	5	4	0	1	0.0668
Sweets and sweet baked products	12	8	0	4	0.0588
Butter and cream	5	0	0	5	0.0754
Spreads and dressings	3	0	1	2	0.0331
Milk	5	2	0	3	0.0735
Soy and other milks	1	1	0	0	-
Soda	1	1	0	0	-
Tea	7	2	0	5	0.2113
Coffee	10	2	6	2	0.1963
Alcohol	17	4	5	8	0.2493

(1) Number of metabolites significantly associated to each food group in the discovery (Bonferroni; $P<7.60\text{x}10^{-6}$) and the discordant MZ twin samples ($P<0.05$), and following fixed effects meta-analysis of results from both groups (Bonferroni; $P<7.60\text{x}10^{-6}$).

(2) Metabolites with multiple associated foods were included in a backwards stepwise regression (cut-off threshold $P<0.05$), those associations not passing the significance threshold from the discovery analysis ($P<7.60\text{x}10^{-6}$) were removed.

(3) Food groups associated with multiple metabolites were included in a backwards stepwise regression, metabolites not passing the cut-off threshold of $P<0.05$ were removed.

Figure 5-2. Metabolites associated to food groups in the training dataset

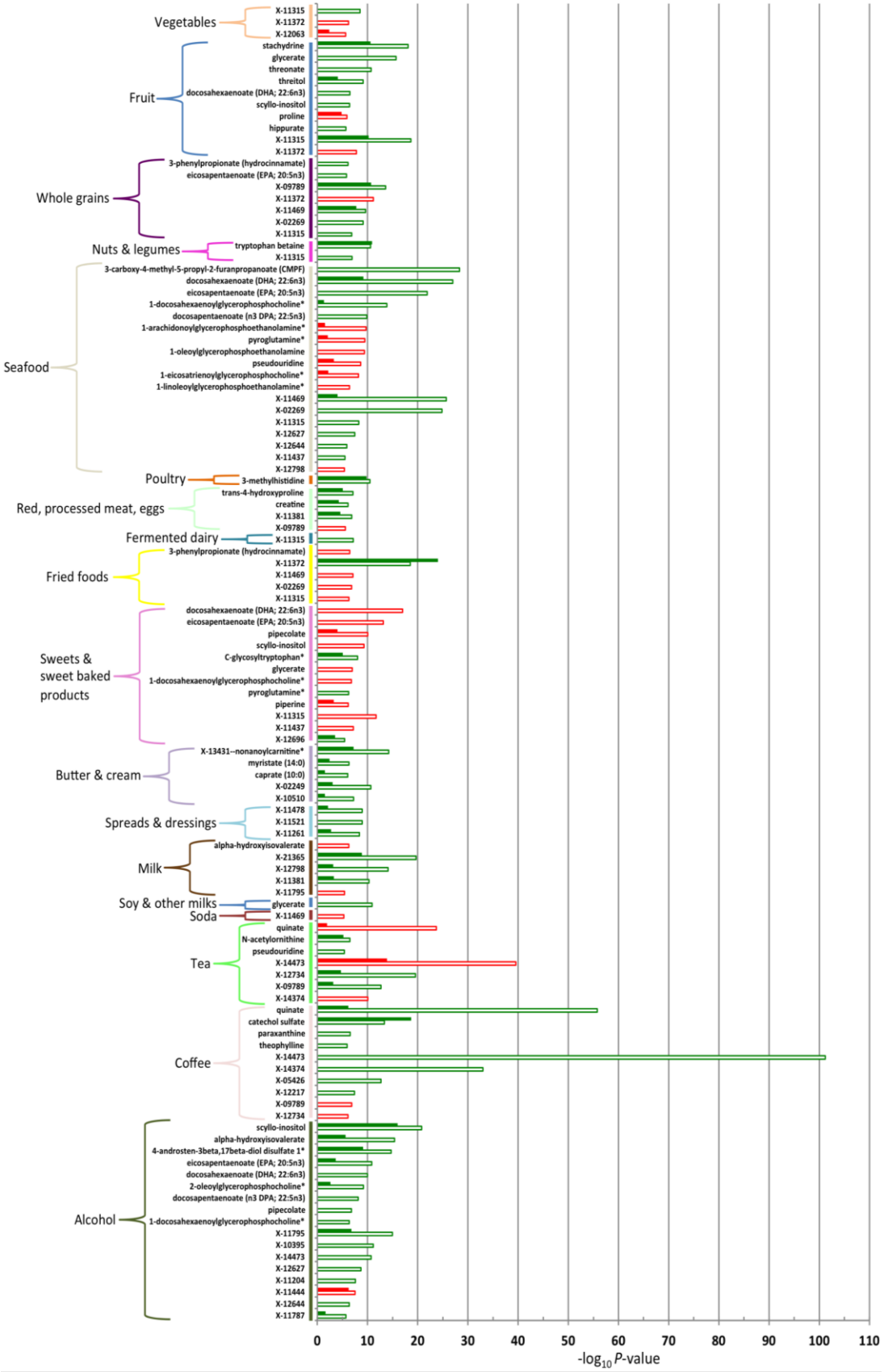


Figure 5-2 shows the metabolites significantly associated with intakes of each food group in the training dataset. Bars with a white centre represent the $-\log_{10}$ *P*-value from the discovery analysis. Solid bars indicate the $-\log_{10}$ *P*-value from the multivariate regression with the final metabolites included in the scores following data reduction. *Green* bars indicate a positive association and *red* bars indicate a negative association. Metabolites whose biochemical identity has not yet been identified are denoted by 'X-' followed by 5 digits.

Table 5-5. List of metabolites associated with a single food group following the discovery analysis

Food group	Metabolite	Superpathway	Subpathway
Vegetables	↓ X-12063	Unknown	
Fruit	↑ stachydrine	Xenobiotics	Food component, Plant
	↑ threonate	Cofactors and vitamins	Ascorbate and aldarate metabolism
	↑ threitol	Carbohydrate	Nucleotide sugars, pentose metabolism
	↓ proline	Amino acid	Urea cycle; arginine-, proline-, metabolism
	↑ hippurate	Xenobiotics	Benzoate metabolism
Nuts and legumes	↑ tryptophan betaine	Amino acid	Tryptophan metabolism
Seafood	↑ 3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF)	Lipid	Fatty acid, dicarboxylate
	↓ 1-arachidonoylglycerophosphoethanolamine*	Lipid	Lysolipid
	↓ 1-oleoylglycerophosphoethanolamine	Lipid	Lysolipid
	↓ 1-eicosatrienoylglycerophosphocholine*	Lipid	Lysolipid
	↓ 1-linoleoylglycerophosphoethanolamine*	Lipid	Lysolipid
White meat	↑ 3-methylhistidine	Amino acid	Histidine metabolism
Red, processed meat and eggs	↑ trans-4-hydroxyproline	Amino acid	Urea cycle; arginine-, proline-, metabolism
	↑ creatine	Amino acid	Creatine metabolism
	↑ C-glycosyltryptophan*	Amino acid	Tryptophan metabolism
	↓ piperine	Xenobiotics	Food component, Plant
	↑ X-12696	Unknown	
Butter and cream	↑ nonanoylcarnitine*	Lipid	Carnitine metabolism
	↑ myristate (14:0)	Lipid	Long chain fatty acid
	↑ caprate (10:0)	Lipid	Medium chain fatty acid
	↑ X-02249	Unknown	
	↑ X-10510	Unknown	
Spreads and dressings	↑ X-11478	Unknown	
	↑ X-11521	Unknown	
	↑ X-11261	Unknown	

Table 5-5. List of metabolites associated with a single food group following the discovery analysis

Food group	Metabolite	Superpathway	Subpathway
Milk	↑ X-21365 (trimethyl-N-aminovalerate)	Unknown	
Tea	↑ N-acetylornithine	Amino acid	Urea cycle; arginine-, proline-, metabolism
Coffee	↑ catechol sulfate	Xenobiotics	Benzoate metabolism
	↑ paraxanthine	Xenobiotics	Xanthine metabolism
	↑ theophylline	Xenobiotics	Xanthine metabolism
	↑ X-05426	Unknown	
	↑ X-12217	Unknown	
Alcohol	↑ 4-androsten-3beta,17beta-diol disulfate 1*	Lipid	Sterol, Steroid
	↑ 2-oleoylglycerophosphocholine*	Lipid	Lysolipid
	↑ X-10395	Unknown	
	↑ X-11204	Unknown	
	↓ X-11444	Unknown	
	↓ X-11787	Unknown	

5.3.2 Overall associations with food group intakes following data reduction

Each food group was associated with more than one metabolite (73 associations overall) which were potentially correlated, therefore a backward stepwise linear regression was performed to remove these using a significance threshold of $P<0.05$. **Table 5-4** summarises the results of this analysis, the detailed results can be found in **Appendix D Table 3**. For alcohol, seafood and coffee intakes, at least one metabolite was significant in the first model but in the opposite direction to the results of the fixed effects meta-analysis (**Appendix D Table 3**, metabolite name and result highlighted in red), therefore a second model was undertaken using only the significant metabolites that were in the same direction. The names and weightings (standardized betas) of the final metabolites used to produce the metabolite scores are highlighted in bold in **Appendix D Table 3**. The metabolites included in the final score are indicated by the solid bars in **Figure 5-2** which represent the $-\log_{10}$ *P*-value from the multivariate regression including all final metabolites in the food score (the R^2 for these regression models can be found in **Table 5-4**).

Complete results for each food group against each final metabolite score can be found in (**Appendix D Table 4**). The associations between the metabolite scores and food group intake were stronger than the top single metabolite association for all food groups in the training dataset. All top metabolites and score associations except for vegetables and spreads and dressings (continuous and weighted) passed the significance threshold for the discovery analysis ($P<7.60\text{x}10^{-6}$) in the test group.

5.3.3 ROC analysis of metabolite food scores in the test group

Figure 5-3 shows the ROC curves for each of the metabolite scores ability to predict high and low consumers based on tertiles (intakes summarised in **Appendix D Table 5**). The results for the top metabolite and top performing metabolite score are presented in **Table 5-6** (results for all scores can be found in **Appendix D Table 6**). Overall, all scores were above the line of no-discrimination (AUC>0.5), indicating the metabolite scores performed better than a random guess at identifying low and high food group consumers. Metabolite scores for alcohol performed moderately (AUC>0.8) and metabolite scores for fruit, whole grains, seafood, fried foods, tea and coffee performed reasonably (AUC>0.7) at predicting high and low consumers. When comparing the combined metabolite scores against the top single metabolite from each score the weighted method performed significantly better ($P<0.05$) and the best for fruit,

seafood, red meat, sweet and sweet baked products, butter and cream and alcohol intakes. For the rest of the groups the scores did not significantly improve the prediction. For foods associated to only a single metabolite (fried foods, vegetables, nuts and legumes, and white meat), dividing the group into quartiles did not significantly change the predictive ability.

5.3.4 Metabolite score associations with food preferences in the test group

We previously showed that food preference patterns in our twin population are correlated to metabolite signatures in the blood, despite metabolomics profiles being measured seven years prior (Pallister et al., 2015). This finding suggests that food preference patterns are quite stable through time. Here I have used the food preference questionnaire to further validate the metabolite scores in the test group. All score associations passed the Bonferroni cut-off ($P<3.33x10^{-3}$) except for spreads and dressings. The top results from this analysis can be found in **Table 5-7**. The full results are in **Appendix D Table 7**. For those foods only associated with 1 metabolite, vegetables and white meat were not significantly associated with food liking, which supports our results from the ROC analysis. These results provide further evidence that food liking has a strong impact on food intakes.

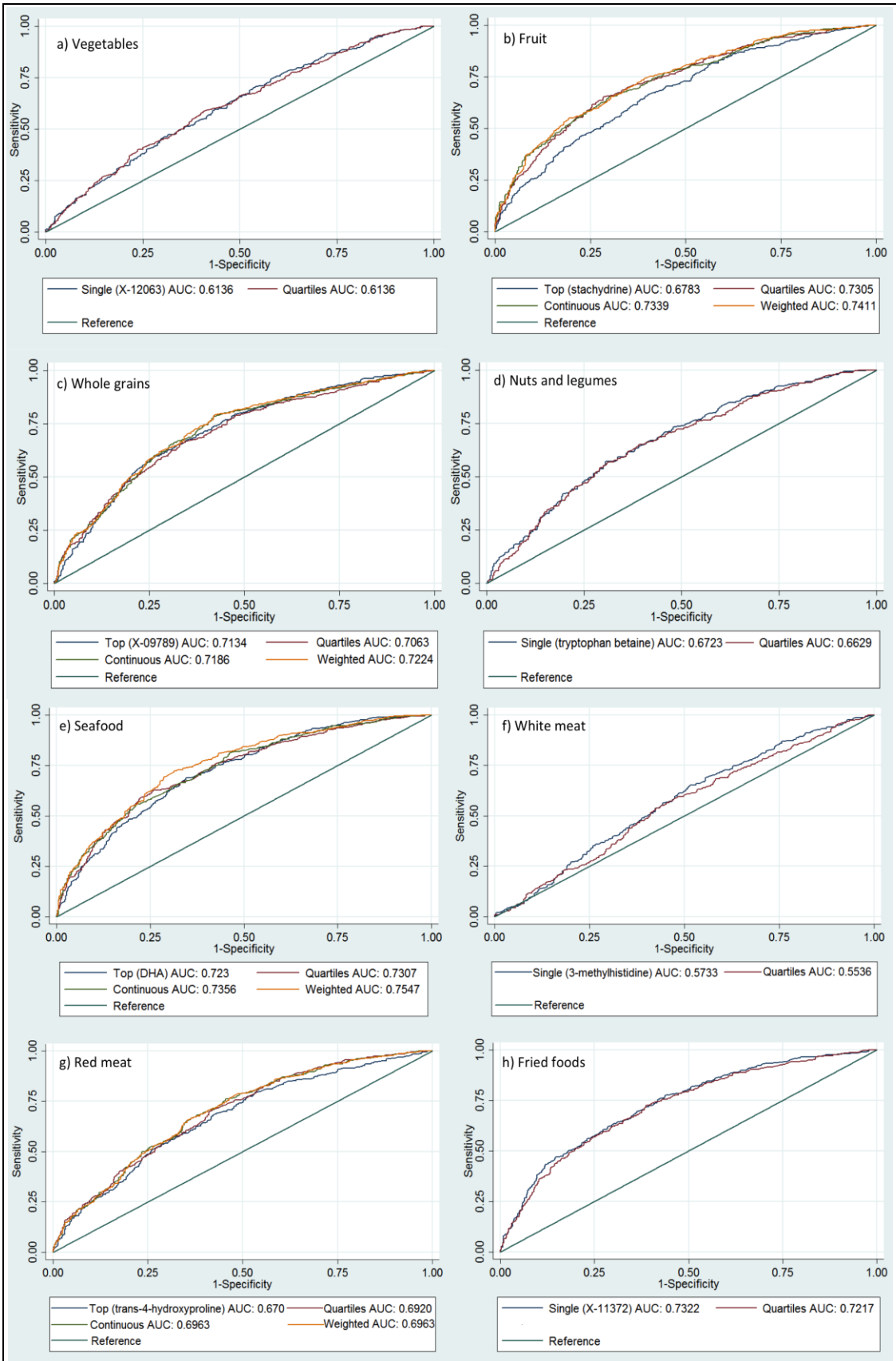
5.3.5 Metabolite scores results summary

i) Summarising multiple metabolites into a score improved the strength of associations between food intakes and metabolite profiles.

ii) Differences in the overall predictive performance of scoring methods were negligible for most food groups, though the weighted method generally performed the best.

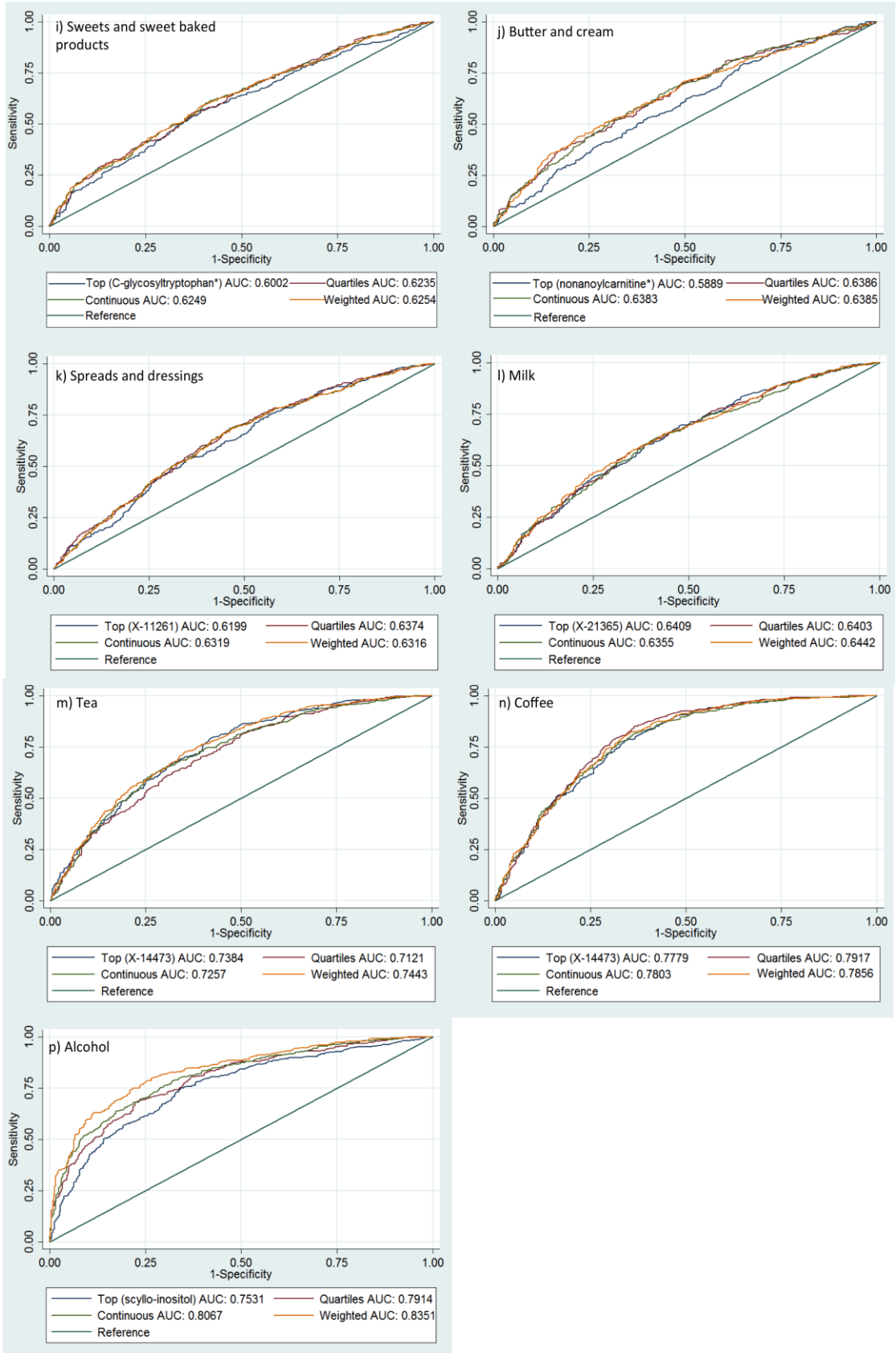
iii) Based on the results from the ROC analysis and associations with food preferences, scores for vegetable, white meat and spreads and dressings did not appear useful.

Figure 5-3. ROC curves for each diet metabolite score



Each food group metabolite scoring method (quartiles, continuous and weighted) and the top associated metabolite were each fitted into a logistic regression model (adjusted for age, BMI, batch and family relatedness) to classify the top (1, positive outcome) and bottom (0, negative outcome) tertiles of food group intake. The equality of the receiver operating characteristic area (AUC) is shown for each scoring method and the top associated metabolite.

Figure 5-3. Continued…ROC curves for each diet metabolite score



Each food group metabolite scoring method (quartiles, continuous and weighted) and the top associated metabolite were each fitted into a logistic regression model (adjusted for age, BMI, batch and family relatedness) to classify the top (1, positive outcome) and bottom (0, negative outcome) tertiles of food group intake. The equality of the receiver operating characteristic area (AUC) is shown for each scoring method and the top associated metabolite.

Table 5-6. Results of ROC analysis [(1)]

						Versus top metabolite	
Food group	**Score**	**Sensitivity**	**Specificity**	**Correctly classified**	**AUC**	χ^2	*P*
Vegetables [(2)]	Top: X-12063	59.74%	55.47%	57.60%	0.6136 [0.5802, 0.6470]		
Fruit	Top: X-11315	65.59%	60.34%	62.98%	0.6783 [0.6479, 0.7087]		
	Weighted	66.84%	67.24%	67.04%	0.7411 [0.7131, 0.7690]	29.77	<0.0001
Whole grains	Top: X-09789	65.25%	67.07%	66.15%	0.7134 [0.6841, 0.7428]		
	Weighted	67.63%	65.81%	66.72%	0.7224 [0.6934, 0.7514]	1.37	0.2425
Nuts and legumes [(2)]	Top: tryptophan betaine	64.25%	60.70%	62.47%	0.6723 [0.6366, 0.7080]		
Seafood	Top: DHA	64.79%	67.46%	66.13%	0.7230 [0.6943, 0.7516]		
	Weighted	69.57%	70.68%	70.13%	0.7547 [0.7274, 0.7821]	11.88	0.0006
White meat [(2)]	Top: 3-methylhistidine	77.78%	32.44%	56.86%	0.5733 [0.5371, 0.6096]		
Red meat	Top: trans-4-hydroxyproline	58.94%	65.37%	62.16%	0.6697 [0.6390, 0.7003]		
	Weighted	63.61%	65.53%	64.57%	0.6963 [0.6666, 0.7260]	9.36	0.0022
Fried foods [(2)]	Top: X-11372	65.30%	66.32%	65.81%	0.7322 [0.7038, 0.7606]		
Sweets and sweet baked products	Top: C-glycosyltryptophan*	58.67%	56.51%	57.59%	0.6002 [0.5679, 0.6325]		
	Weighted	60.10%	58.97%	59.54%	0.6254 [0.5937, 0.6571]	5.07	0.0243
Butter and cream	Top: nonanoylcarnitine*	71.88%	39.55%	56.84%	0.5889 [0.5529, 0.6249]		
	Weighted	59.15%	59.90%	59.52%	0.6385 [0.6036, 0.6734]	11.39	0.0007
Spreads and dressings	Top: X-11261	59.08%	57.48%	58.28%	0.6199 [0.5880, 0.6518]		
	Quartiles	60.27%	60.10%	60.19%	0.6374 [0.6059, 0.6689]	5.42	0.0199
Milk	Top: X-21365 [trimethyl-N-aminovalerate]	60.17%	60.65%	60.41%	0.6409 [0.6096, 0.6723]		
	Continuous	60.68%	61.33%	61.00%	0.6355 [0.6040, 0.6669]	0.22	0.6402
Tea	Top: X-14473	59.28%	73.51%	67.04%	0.7384 [0.7078, 0.7691]		
	Quartiles	61.63%	64.63%	63.13%	0.7121 [0.6803, 0.7439]	2.68	0.1016
Coffee	Top: X-14473	78.21%	65.38%	72.42%	0.7779 [0.7483, 0.8074]		
	Quartiles	72.06%	76.36%	74.21%	0.7917 [0.7627, 0.8208]	3.08	0.079
Alcohol	Top: scyllo-inositol	75.75%	63.32%	70.02%	0.7531 [0.7230, 0.7832]		
	Weighted	74.70%	76.40%	75.55%	0.8351 [0.8104, 0.8598]	38.99	<0.0001

ROC, receiver operating characteristic; AUC area under the receiver operating characteristic curve

(1) Each food group metabolite scoring method (quartiles, continuous and weighted) and the top associated metabolite were each fitted into a logistic regression model (adjusted for age, BMI, batch and family relatedness) to classify the top (1, positive outcome) and bottom (0, negative outcome) tertiles of food group intake. The equality of the receiver operating characteristic area (AUC) for each scoring method was tested against the ROC area for the top associated metabolite.

(2) Only one metabolite marker

Table 5-7. Top and metabolite score associations with food preferences in the test group [1]

Food group	Top metabolite	Beta (SE)	*P*	Best scoring method	Beta (SE)	*P*
Vegetables [2]	X-12063	-0.062 (0.059)	*NS*			
Fruit	X-11315	0.203 (0.045)	7.40×10^{-6}	Weighted	0.077 (0.017)	1.12×10^{-5}
Whole grains	X-09789	0.287 (0.054)	1.37×10^{-7}	Weighted	0.045 (0.012)	1.70×10^{-4}
Nuts and legumes [2]	tryptophan betaine	0.271 (0.045)	3.04×10^{-9}			
Seafood	DHA	0.333 (0.035)	6.85×10^{-20}	Weighted	0.173 (0.014)	1.51×10^{-30}
White meat [2]	3-methylhistidine	0.083 (0.037)	2.45×10^{-2}*			
Red meat	trans-4-hydroxyproline	0.221 (0.046)	2.29×10^{-6}	Weighted	0.081 (0.009)	7.71×10^{-19}
Fried foods [2]	X-11372	0.344 (0.046)	2.69×10^{-13}			
Sweets and sweet baked products	C-glycosyltryptophan	0.004 (0.002)	1.09×10^{-2}*	Weighted	0.002 (0.000)	2.24×10^{-7}
Butter and creams	nonanoylcarnitine	0.023 (0.033)	4.79×10^{-1}*	Weighted	0.022 (0.007)	2.61×10^{-3}
Spreads and dressings	X-11261	0.105 (0.041)	1.15×10^{-2}*	Quartiles	0.207 (0.089)	2.06×10^{-2}*
Milk	X-21365 [trimethyl-N-aminovalerate]	0.167 (0.041)	5.71×10^{-5}	Continuous	0.352 (0.097)	3.15×10^{-4}
Tea	X-14473	-0.217 (0.031)	7.83×10^{-12}	Quartiles	0.624 (0.082)	8.88×10^{-14}
Coffee	X-14473	0.241 (0.055)	1.36×10^{-5}	Quartiles	0.689 (0.101)	1.98×10^{-11}
Alcohol	scyllo-inositol	0.322 (0.042)	1.01×10^{-13}	Weighted	0.197 (0.018)	4.08×10^{-27}

NS, not significant; *, not statistically significant ($P < 3.33 \times 10^{-3}$)

(1) The top metabolite associated with each food group and the best performing scoring method (if multiple metabolites associated) from the receiver operating characteristic curve analysis were using to predict preferences for food groups.

(2) Only one metabolite was associated with the food group.

5.3.6 Metabolite score associations with future incidence of MetS

Metabolite scores for butter and creams (OR[SE]: 0.602[0.119]; *P*=0.010), seafood (0.513[0.115]; $P=2.85\times10^{-3}$), and whole grains (0.325[0.095]; $P=1.15\times10^{-4}$) were associated with reduced odds of future MetS. Reported intakes for the same food groups were not significantly predictive of MetS, though effect directions were the same, which is suggestive that the metabolite scores were able to capture an effect more strongly than the questionnaires. Each significant metabolite score was predictive of at least one feature of MetS. Lower metabolite scores were predictive of reduced HDL-cholesterol for all food groups (**Table 5-8**). Lower metabolite scores for whole grains were mildly predictive of elevated triglycerides. There were no associations between metabolite scores and elevated WC, high blood pressure or impaired fasting plasma glucose.

Table 5-8. Metabolite score predictions of criteria categories of the metabolic syndrome

Criteria	Butter and creams		Seafood		Whole grains	
	OR[SE]	*P*	OR[SE]	*P*	OR[SE]	*P*
Elevated triglycerides		*NS*		*NS*	0.639[0.123]	0.020
Reduced HDL-cholesterol	0.347[0.125]	3.38×10^{-3}	0.413[0.123]	3.05×10^{-3}	0.367[0.099]	2.12×10^{-4}

NS, not significant (*P*<0.05); MetS, metabolic syndrome; HDL, high density lipoprotein cholesterol
A logistic regression (0, no risk; 1, risk) was performed in 414 twins using each food metabolite score (residual adjusted for age, BMI and batch at time of metabolomics) to predict the 5 clinical MetS criteria taken ≥ 5 years after metabolomics profiling.

5.4 Discussion

The primary aims of this chapter were to identify markers of long-term reported intakes of general food groups, to create biomarker scores using multiple methods and evaluate them, including testing them on the future development of MetS. Through the work I completed in this chapter I came to the following conclusions: i) summarising multiple metabolites into a score improved the strength of associations between food intakes and metabolite profiles; ii) differences in the overall predictive performance of scoring methods were negligible for most food groups, though the weighted method generally performed the best; iii) based on the results from the ROC analysis and associations with food preferences, I would not use scores for vegetable, white meat and spreads and dressings for further analysis. Nearly all of the metabolite associations identified and confirmed in discordant MZ twins have been identified in Chapter 4 as markers of reported intakes of more specific food items, I have therefore not highlighted these in the discussion. Moreover, I found that lower biomarker scores, though not reported intakes, associated to consumption of butter and creams, seafood and whole grain products associated to a higher risk of MetS later in life.

Few epidemiological studies and no twin studies to date have examined reported food intakes against metabolomics profiles to create biomarker scores and evaluated them. One study of African Americans examined the impact of alcohol consumption related metabolite sub-pathways on white blood cell levels and incident CVD (Zheng et al., 2014). In this chapter I used their method of summing quartile ranked metabolites to create a score, though this method did not perform the best overall and appears to lose some information. Moreover, while the authors created scores for different sub-pathways they did not acknowledge that these pathways are not specific to alcohol intake, such as 2-hydroxybutyrate and related branched chain amino acids which are markers of insulin resistance and type 2 diabetes (Menni et al., 2013a), and therefore may not truly connect alcohol consumption to CVD directly. I also used the summation score method that has previously been used to assess adherence to a healthy Nordic diet (Marklund et al., 2014). I found this method simple, though for most food groups did not perform better than the weighted method. I used the weighted method as it has been used to study markers of oxidative stress and their influence on incident, sporadic colorectal adenomas (Dash et al., 2013), though they did not directly test the difference in accuracy between the equal weight versus weighting methods, the authors did find the methods to predict colorectal adenomas

similarly. An issue with equally weighting scores is that they may be less biologically meaningful than weighting. However, the weights used here may not be applicable to other populations. Although through splitting our group into a training and test set I found the weighting method still tended to perform better. Single metabolites did not perform well compared with combining metabolites into scores, supporting the use of biomarker scores for future study.

All scores performed better than a random guess at identifying low and high food group consumers. Generally, scores including more metabolites performed better. Scores that performed with moderate accuracy (AUC>0.7) included those for alcohol, fruit, whole grains, seafood, fried foods, tea and coffee. Suggesting that these metabolites should be investigated more thoroughly. Previous metabolomics studies have identified strong associations with reported coffee, fruit, alcohol and seafood intakes (Guertin et al., 2014b, Guertin et al., 2015, Zheng et al., 2014), many of which I confirmed in this and the previous chapter. It is interesting, though encouraging, that the marker for fried food intake performed reasonably well as reporting of fried food intake may be more prone to underreporting. It is difficult to draw conclusions for metabolite scores that did not perform as well (AUC<0.7). However, the potential markers may simply not be good markers, though it also should be taken into account that ranking of the intakes for some foods may not be precise. To help resolve these issues controlled feeding studies should be undertaken, moreover applying different or new metabolomics platforms with more metabolites may yield markers with greater potential.

An important issue with creating scores is that many of the metabolites associated to food intakes appear to lack specificity to a particular food. Moreover, in our study many metabolites were associated to multiple foods that may also be a result of correlations in reporting and dietary patterns. I attempted to ameliorate this issue by running a multivariate linear regression including all foods associated to each applicable metabolite and used the stringent significance threshold from the discovery analysis. However, by doing this I may have eliminated important markers. Moreover, future studies may find that working with correlations in intakes and identifying markers of healthy versus unhealthy dietary patterns (which tend to cluster together) may be more useful for public health research (and easier for translation) than examining single foods alone.

I found that lower biomarker scores (though not reported intakes) for butter and creams, seafood and whole grains associated to a higher risk of MetS later in life. However, only 25 individuals subsequently developed MetS at a later time, therefore strong conclusions cannot be

drawn from this small sample size. It is interesting that metabolites associated to butter and cream intake were lower in individuals who later developed MetS, which seems conflicting. Moreover, higher scores on the butter and cream metabolite score were associated with higher levels of HDL-cholesterol. Although large population studies have supported these findings, a recent study on 15,105 Brazilian adults identified inverse associations between total dairy and high fat dairy consumption and later development of the MetS which appeared to be mediated by dairy saturated fat consumption (Drehmer et al., 2016). In a large population of adults from 2 US cohorts (the Health Professionals Follow-Up Study: 51,529 men; the Nurses' Health Study: 121,700 women), classic blood biomarkers of dairy fat consumption, *trans* palmitoleate, pentadecanoic and heptadecanoic acids were not associated with incidence of stroke (Yakoob et al., 2014). In another study of 659 adults from the triethnic multicenter Insulin Resistance Atherosclerosis Study, circulating pentadecanoic acid was inversely associated with incident type 2 diabetes, a clinical measure of insulin resistance and β-cell dysfunction (Santaren et al., 2014). I also found that individuals with lower whole grain metabolite scores were more likely to develop MetS. All of the metabolites forming the whole grain score are unknown therefore not much information can be derived from them at this time. Though it has been found that whole grain consumption appears to have a protective effect on MetS (Sahyoun et al., 2006) and cardiometabolic traits (Giacco et al., 2010). Using a biomarker approach, increased plasma alkylresorcinols (AR) ratio C17:0/C21:0, a biomarker of whole grain rye intake, was associated with an improved blood lipid profile following the consumption of a healthy Nordic diet in subjects with MetS (which includes high intakes of rye bread) (Magnusdottir et al., 2014), however total AR which is more representative of total whole grain wheat and rye intake was not associated with an improved lipid profile. I identified a particularly strong association between having lower scores on the whole grain metabolite score and having lower HDL-cholesterol, though a strong relationship between HDL-cholesterol and whole grain consumption has not been identified (Hollaender et al., 2015). Higher scores on the seafood biomarker score were also found to associate with a reduced MetS risk and higher HDL cholesterol. A systematic review of 7 studies found 4 studies (one follow-up, three cross-sectional) to suggest increased fish intake is protective of MetS (Torris et al., 2014). The long chain omega-3 fatty acids found in fatty fish, and also included in the biomarker score I used, are thought to be protective through improving vascular function and reducing inflammatory processes (Tousoulis et al., 2014). Supporting this, a recently published prospective study of

4356 American young adults found that increasing reported intakes of non-fried fish and dietary long-chain omega-3 fatty acids were associated inversely with MetS at 25 years follow-up (Kim et al., 2016). Biomarker studies are scarce on the relationship between blood long-chain omega-3 PUFAs and MetS risk. One study of chronic kidney disease patients found a pattern of high *n*-3 PUFAs in serum were not consistently associated with MetS (Huang et al., 2014). In another study serum DHA was associated with reduced cardiometabolic risk (Song et al., 2014).

There are a number of limitations and factors that must be considered. The Metabolon method does not determine the true blood concentration of the metabolites therefore I could not identify accurate cut-off points for what constitutes high and low intakes or risk levels. For chocolate, refined grains, fermented dairy, soy and other milk, and soda intakes no metabolite biomarkers were identified or did not pass data reduction and could not be studied further. By evaluating the metabolite scores against FFQ data I likely did not capture the true discriminatory ability of the scores due to the nature of dietary self-reporting. Through using these methods I was able to compare the utility of different scoring methods. Future feeding studies will be better equipped to assess biomarker utilization more accurately. The classification of foods was done *a priori* based on foods derived from similar sources (e.g. plants, animals, processing), which confer similar health benefits (e.g. whole grain versus refined grain), food usage and cooking. I may have chosen foods to be in a group that are not well correlated to intakes of the rest of the foods, therefore one food may be more strongly related to a metabolite, whereas another food included in the group has no relationship. Future work should aim to unravel the precise mechanism for these associations in order to discern which foods contribute to the same metabolite associations. Only women were used for this study, therefore the results may not be applicable to men. Many markers are unidentified at this time therefore I could not confirm they may be biologically related to food intakes, however I used stringent cut-offs, discordant MZ twins and data reduction methods to identify the best metabolite markers possible in our population. Importantly, the metabolite scores were tested in a group independent from the discovery group.

Chapter 6 Metabolomic associates of microbial diversity, their modulation by diet and relationship to the metabolic syndrome

In this chapter I analysed gut microbiome Shannon diversity (Shannon) for its association with blood metabolites from the Metabolon platform, I then examined whether these microbe-related metabolites were associated with reported food intakes and finally, I explored longitudinal changes of the top diversity-associated metabolite and its association with metabolic syndrome risk and its components.

6.1 Introduction

The microbiome refers to the collective genomes of the microorganisms within an environmental niche. Within the human gut, microbiota undertake important processes allowing for potential host-microbe interactions such as vitamin synthesis, hormone production, metabolism of food components and interactions with host innate immunity. The number of different organisms present within the human gut (referred to as 'richness' or 'diversity') is associated with the increased abundance of beneficial bacteria and is emerging as an important indicator of health. Richness represents the total gene count, while diversity additionally accounts for the rarity of those microbial genes present within the whole dataset. Reduced alpha-diversity (intra-individual diversity) has been coupled with dysbiosis (microbial imbalance) in inflammatory bowel disease patients (Manichanh et al., 2006, Michail et al., 2012, Scher et al., 2015), though this has not been highlighted in studies of obesity (Turnbaugh et al., 2009) or metabolic health (Le Chatelier et al., 2013). Although mechanisms are unclear microbial imbalance occurs when beneficial bacterial growth is compromised and a few pathogenic taxa dominate and overtake the metabolic potential of the microbiome with possible deleterious effects. Recent studies have shown reduced diversity is present in disease states and metabolic conditions such as auto-immune diseases, including inflammatory bowel disease (Michail et al., 2012, Ott et al., 2004, Manichanh et al., 2006) and in patients with psoriatic arthritis and inflammatory bowel disease (Scher et al., 2015)**,** and metabolic derangements, including obesity (Turnbaugh et al., 2009) and MetS-related phenotypes (Le Chatelier et al., 2013). Diet has been shown to contribute to gut microbiome richness (Cotillard et al., 2013), in particular dietary fibre promotes stability in

richness in humans (Tap et al., 2015) and this stability has recently been shown to be transferred to future generations in mice (Sonnenburg et al., 2016), highlighting the important potential of good nutrition in humans.

Microbes play a distinct role in human metabolism by transforming food- and host-derived metabolites, such as bile acids, polyphenols and fibre. These products may in turn influence disease development by interacting with human host physiology. Studies are now attempting to untangle these interactions through measuring the chemical profile (metabolome) in biofluids (urine and blood), faeces, tissues and organs in collaboration with microbiome analysis. Comparing conventional versus germ-free mice, conventional mice exhibited elevated blood levels of indole-containing compounds (e.g. indoxyl sulfate and indole-3-propionic acid), serotonin, sulfated compounds (e.g. phenyl and *p*-cresol sulfate), and glycine-conjugated compounds (hippuric acid, cinnamoylglycine and phenylpropionylglycine), showing the major contribution of the gut microbiome to metabolism (Wikoff et al., 2009). In another study, urinary excretion of hippuric acid, a metabolite derived from high polyphenol foods (Gonthier et al., 2003, Walsh et al., 2007), has been found to discriminate well between obese and normal weight controls (Calvani et al., 2010). Many of the metabolites implicated are food-derived components. In the two previous chapters I found many food associations with microbial co-metabolites, therefore merging microbiome and metabolomics approaches with dietary studies is the logical next step in improving our understanding of the interplay between diet, the microbiome and metabolic disease.

Diet is thought to be the most important avenue for modulating the gut microbial composition and its metabolic outputs, despite this there are relatively few good human studies. It has been shown that daily consumption of 40 g of dark chocolate for 2 weeks altered urinary output of gut microbial metabolites, increasing hippurate and methylamines, and reducing p-cresol sulfate (Martin et al., 2009). A recent randomized controlled pilot found consuming 30 g/d of heat-stabilized rice bran for 28 days increased abundance of 8 operational taxonomic units (OTUs) from the genera *Methanobrevibacter*, *Paraprevotella*, *Ruminococcus*, *Dialister*, *Anaerostipes* and *Barnesiella*, and 3 from *Bifidobacterium* and *Ruminococcus*, moreover elevated levels of secondary bile acids (deoxycholic and lithocholic) and metabolites derived from plants and microbial modifications of plant phenolics (benzoic, hydrocinnamic and phosphoric acids, and inositol monophosphate) were observed in the faecal metabolome (Sheflin et al., 2015). The microbial metabolism of choline and L-carnitine, mainly found in meat,

fish and eggs, to trimethylamine (TMA) and then oxidized in the liver to trimethylamine-N-oxide (TMAO) has been implicated in why some red-meat consumers develop atherosclerosis (Koeth et al., 2013, Wang et al., 2011). Specifically, TMAO has been suggested to encourage the upregulation of macrophage scavenger receptors and through this enhance forward cholesterol transport (Bremer, 1983). Though in mice with intact intestinal microbiota and also with higher TMA and TMAO blood concentrations, choline and L-carnitine feeding each suppressed reverse cholesterol transport (Koeth et al., 2013). Interestingly, suppression of the gut microbiota removed the effect.

The metabolomic signature of gut microbial alpha-diversity has yet to be examined on a large scale, and could provide useful candidate markers of a metabolically fit microbiome. Therefore for the current chapter, I aimed to:

i) Identify the blood metabolites correlated with gut microbiome diversity.
ii) Examine the impact of food intake on these metabolites.
iii) Examine if longitudinal changes in these metabolites are predictive of both future metabolic syndrome and its components.

6.2 Materials and methods

Data relevant to this chapter included the reported intakes of 20 food groups as described in the previous chapter (**Section 5.3.1**), microbiome data and blood metabolomics data (details **Section 3.1.6.1**). I used all 292 known metabolites measured by the Metabolon platform. Twins who completed FFQs collected between 1995 and 2001 and in 2007 were used in the discovery analysis (*n*=1529). New FFQ data were collected between 2014 and 2015, 484 additional twins had microbiome, metabolomics data and completed FFQs during this time, 420 of these twins who had no co-twin in the discovery sample were used as a validation sample. The remaining 64 twins who had co-twins in the discovery sample were used for the subsequent microbiome and MetS analyses.

Table 6.1 provides the study population characteristics and subject numbers.

Table 6-1. Study population characteristics for the whole, discovery and validation samples

	Whole	FFQ<2014 (Discovery)	FFQ≥2014 (including validation)
	Mean (SD)	Mean (SD)	Mean (SD)
N	2013	1529	484
Age (y)	57.2 (10.6)	57.7 (10.6)	55.4 (10.4)
BMI	26.0 (4.6)	26.1 (4.6)	25.9 (4.6)
Sex (M:F)	113:1909	0:1535	113:374
MZ:DZ pairs	408:392		
Singletons	421		
Food groups (servings/week)			
Vegetables	34.9 (16.5)	34.8 (15.4)	35.1 (19.6)
Fruit	21.4 (12.7)	21.9 (12.4)	19.5 (13.5)
Whole grains	10.3 (8.0)	10.7 (8.1)	9.2 (7.4)
Refined grains	8.6 (7.6)	9.0 (7.9)	7.2 (6.3)
Nuts and legumes	7.9 (5.8)	7.6 (5.2)	8.8 (7.2)
Seafood	2.4 (2.0)	2.5 (2.0)	2.3 (2.1)
White meat	1.9 (1.3)	1.9 (1.3)	1.9 (1.3)
Red meat	6.9 (4.1)	6.8 (3.9)	7.4 (4.6)
Fermented dairy	6.2 (5.1)	6.1 (4.7)	6.5 (6.2)
Fried foods	4.5 (3.5)	4.5 (3.3)	4.8 (4.1)
Sweets and sweet baked products	15.2 (13.9)	15.5 (14.0)	14.1 (13.4)
Chocolate	4.0 (5.9)	4.0 (5.8)	3.9 (6.3)
Butter and cream	4.4 (6.5)	4.2 (6.4)	5.0 (6.8)
Spreads and dressings	8.2 (9.0)	8.5 (9.1)	7.2 (8.6)
Milk	3.2 (2.3)	3.4 (2.3)	2.8 (2.2)
Soya and other milk	0.2 (0.8)	0.2 (0.9)	0.2 (0.8)
Soda	1.7 (4.2)	1.8 (4.3)	1.3 (3.5)
Tea	19.3 (13.5)	19.2 (13.5)	19.4 (13.5)
Coffee	9.0 (10.6)	9.1 (10.7)	8.8 (10.5)
Alcohol	6.1 (8.0)	6.2 (7.9)	5.9 (8.4)

6.2.1 Faecal microbiome composition

Faecal samples were collected at follow-up and the composition of the gut microbiome was determined by 16S rRNA gene sequencing carried out at Cornell as previously described (Goodrich et al., 2016). Firstly, the V4 region of the 16S rRNA gene was amplified and then sequenced on Illumina MiSeq. The reads were next compiled to operational taxonomic units (OTUs) (Ley et al., 2006). Quality control was undertaken by Matthew Jackson and Tiphaine Martin at KCL by sample, paired-ends with an overlap of less than 200nt were removed. Chimeric sequences were then removed by *de novo* chimera detection in USEARCH (Edgar et al., 2011). Using Sumaclust within QIIME 1.9.0 *de novo* OTU clustering was undertaken across all reads, reads with a 97% identity threshold were brought together (Jackson et al., 2016a, Caporaso et al., 2010). Log transformation was undertaken on OTU counts, prior to transformation zero counts were given an arbitrary value ($1x10^{-6}$). OTU abundances were then adjusted for technical covariates including sequencing depth, sequencing run, sequencing technician and sample collection method using linear modelling and the residuals obtained. As the residuals were not normally distributed, I completed an inverse normalisation. To determine alpha diversity, the total OTU count table was rarefied to 10000 sequences for each sample 50 times. Per sample, alpha diversity metrics were determined in each of the rarefied tables and the average score for all 50 was considered as the final diversity measure. The primary alpha diversity metric considered was the Shannon Index as this was considered the most robust and is a commonly used metric though findings were confirmed on other parameters including observed OTU counts and Chao1 (richness), and the Simpson diversity index. All alpha diversity indices were standardised to have mean 0 and SD 1.

6.2.2 Classification of the metabolic syndrome

I determined MetS status using the criteria outlined by the International Diabetes Federation and the American Heart Association/National Heart, Lung, and Blood Institute (Alberti et al., 2009). Consult **Section 5.2.4** for details.

Table 6-2 provides the clinical characteristics of the subsample of twins studied.

Table 6-2. Clinical characteristics of the twin subsample to investigate longitudinal hippurate, diversity, diet with the MetS phenotypes and its components

	Londitudinal sample (n=1032)
	Mean (SD)
Age at MetS status (y)	64.2 (7.8)
MetS status (0, no; 1, yes)	906:116
Longitudinal metabolomics baseline to endpoint (y)	10.6 (3.9)
Sex (M:F)	27:1005
BMI (kg/m2)	26.2 (4.5)
Systolic blood pressure (mmHg)	129.0 (15.7)
Diastolic blood pressure (mmHg)	75.5 (9.7)
Glucose (mmol/L)	4.9 (0.6)
Cholesterol (mmol/L)	5.6 (1.0)
HDL-Cholesterol (mmol/L)	1.9 (0.5)
Triglycerides (mmol/L)	1.1 (0.5)

6.2.3 Statistical analysis

Statistical analysis was carried out using Stata version 12. The statistical analysis was undertaken in two parts. First, a marker of microbiome diversity was identified, its relationship to food intake explored and associations with microbiome OTUs/collapsed taxonomies identified. In the second part, the relationship between longitudinal levels of the diversity metabolite marker, diet, diversity and associated OTUs/taxa with the risk of MetS and its components were explored.

6.2.3.1 Part 1: Metabolite associations with diversity, relationship to food intake and associations with microbiome OTUs/taxa

Figure 6-1 shows the analysis pipeline for part 1.

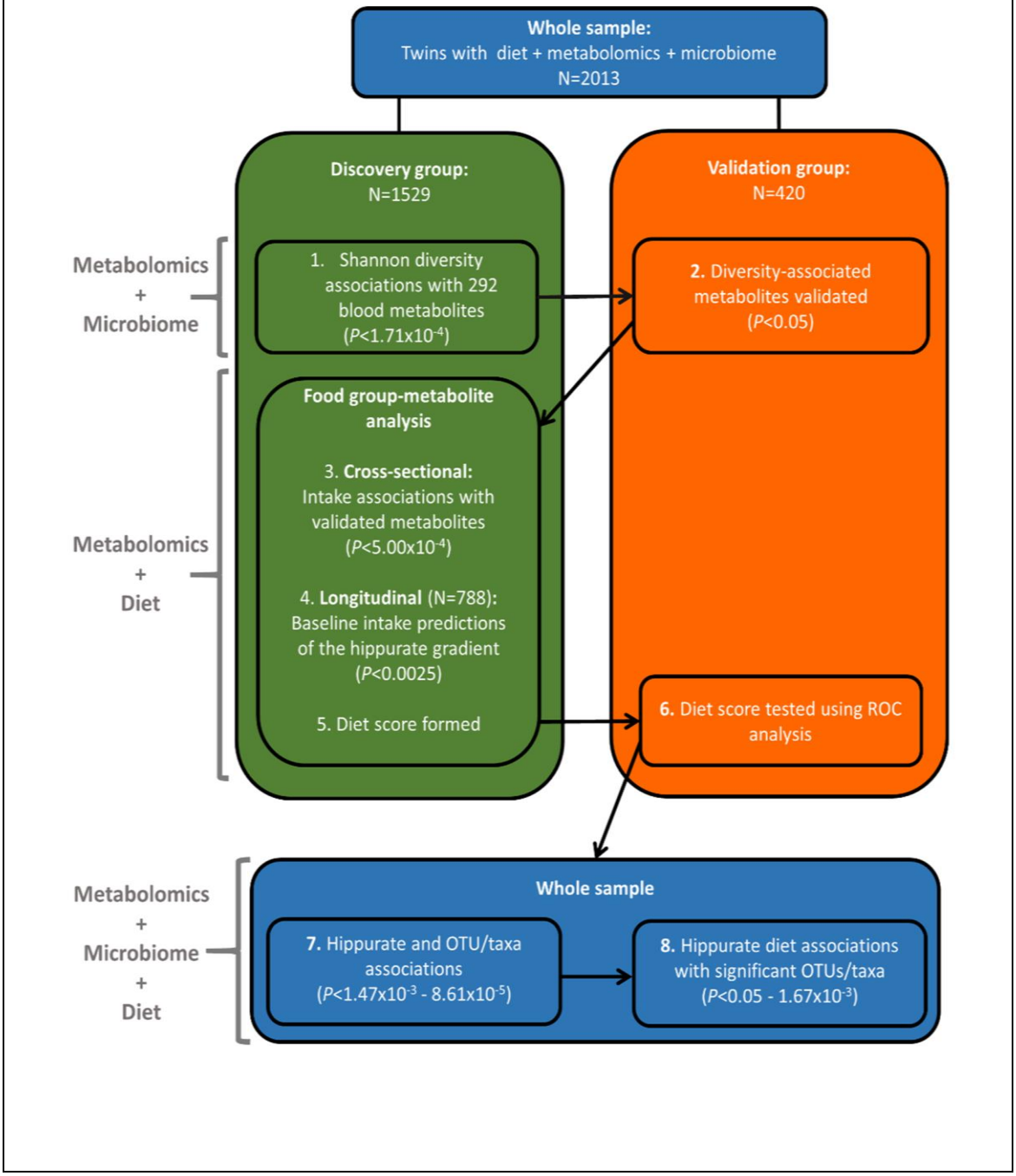
6.2.3.1.1 Microbiome diversity-metabolite associations

I ran a linear regression model with Shannon diversity as a predictor of the metabolite level (292 metabolites, Metabolon platform) in a group of 1529 female twins, adjusted for age, batch, BMI and family relatedness (closest to blood sampling) and multiple testing (Bonferroni: 0.05/292 = 1.71×10^{-4}). Significant metabolites from the discovery sample were then evaluated against Shannon diversity using the same linear regression (additionally including sex as a covariate) in the male (*n*=113) and female validation sample (*n*=420), associations passing the 5% level of significance were considered validated.

6.2.3.1.2 Food intakes associated with diversity-associated metabolites

To identify if any validated diversity-associated metabolites were potentially related to food intake, in the discovery sample I used reported intakes of the 20 food groups to predict levels of the metabolites (Bonferroni: 0.05/[5 metabolites x 20 food groups] = 5.00×10^{-4}), adjusted for the same covariates as above. In a subsample of individuals from the discovery group (*n*=788) longitudinal metabolomics data were available (*n*=705 3 time points, *n*=83 2 time points) as well as reported food group intake at the same time or 5 years before the first blood sample.

Figure 6-1. Flowchart of study design for part 1



I determined trajectories of change in the top diversity-associated metabolite (years after baseline: 2nd time point: 7.4 [range: 2.3-12.8]; 3rd time point: 13.9 [range: 8.3-17.9]) by empirical Bayes predictions (adjusted for age and BMI) which estimates the rate of change in standard deviations/year (Rabe-Hesketh and Skrondal, 2008). Using this method point estimates were calculated and a slope of change determined. Food group intake at baseline was then used to predict the metabolite trajectories (Bonferroni: 0.05/20 food groups = 0.0025).

A predictive score was created from those significantly associated foods according to the direction of association (i.e. positive association: Q1=0, Q2=1, Q3=2, Q4=3; negative association: Q1=3, Q2=2, Q3=1, Q4=0) and summed. In the validation sample, I conducted a logistic regression model (adjusted for covariates, including sex) to test the ability of the top diversity-associated metabolite to identify twins in the top and bottom tertiles of Shannon diversity, as well as the utility of the diet score to identify the top and bottom tertiles of the metabolite in blood.

6.2.3.1.3 Food-microbiome-metabolite axis

To identify associations between the metabolite and the microbiome I combined the discovery and validation samples (*n*=2013). To establish the strong association between the metabolite and diet score with richness (observed OTUs) and additional diversity metrics (Simpson and Chao1), I ran these associations using the same linear regression as for the Shannon diversity discovery analysis in the pooled sample. I then evaluated associations within the microbiome by running a linear regression model using the OTUs and OTUs collapsed at each taxonomic level (phylum, class, order, family, genus) as predictors of the metabolite adjusted for covariates, Shannon diversity and multiple-testing (Bonferroni cut-off; **Table 6-3** shows the significance threshold for the OTUs and each taxonomic level). To determine the total variance in both Shannon diversity and the metabolite accounted for by metabolite-associated OTUs, I included all associated OTUs in a backwards stepwise linear regression using $P<0.05$ as the threshold cut-off, I report the R^2 for each model. Metabolite-associated OTUs/taxa were examined for their association with the diet score (predictor) adjusted for covariates, the metabolite and multiple testing (assigned at each taxonomic level; **Table 6-3**). To investigate if any of the foods forming the score were driving associations, I ran a multivariate regression model including all metabolite-associated foods and the same covariates.

Table 6-3. Statistical significance thresholds for microbiome analysis using Bonferroni correction [(1)]

Level	Hippurate		Hippurate diet [(2)]	
	Number of variables	*P*	Number of variables	*P*
Phylum	34	$1.47x10^{-3}$	1	0.05
Class	**72**	$6.94x10^{-4}$	2	0.025
Order	137	$3.65x10^{-4}$	3	0.017
Family	194	$2.58x10^{-4}$	3	0.017
Genus	382	$1.31x10^{-4}$	7	$7.14x10^{-3}$
OTU	581	$8.61x10^{-5}$	30	$1.67x10^{-3}$

(1) Bonferroni correction was calculated within each level.
(2) Variables significantly associated with hippurate were tested for their association against the hippurate diet score.

6.2.3.1.4 Microbiome-metabolite-diet interactions

I next tested if there was an interaction between the diet score and the microbiome OTU/collapsed taxa associated to both the metabolite and the diet score in the prediction of the blood metabolite levels. I first ran a continuous x continuous interaction model using each OTU/collapsed taxa adjusting for covariates (age, BMI, batch, sex and Shannon diversity). In the second analysis, to determine if total numbers of all associated bacteria were important, I first multiplied abundances of associated microbiome variables by the direction of association with the metabolite (-1, negative, 1, positive) and summed the variables. I then ran the same continuous x continuous interaction model as above using this variable.

6.2.3.2 Part 2: The relationship between longitudinal levels of the diversity metabolite marker, diet, diversity and associated OTUs/taxa with the risk of MetS and its components

Figure 6-2 shows the analysis pipeline for part 2.

6.2.3.2.1 Relationship between microbiome diversity, longitudinal metabolite and the MetS score and its components

Microbiome diversity has been reduced in metabolic diseases, therefore I hypothesized that reduced longitudinal changes in the top metabolite diversity marker would be associated with increased MetS risk. A subsample of 1032 individuals had longitudinal blood metabolite levels (n=533 with 2 time points, n=499 with 3 time points). The time for the longitudinal analysis ranged from 2.4-17.9 years. I evaluated whether the longitudinal levels of the top diversity

metabolite predicted MetS status and scoring or any of its components. I ran a linear regression model using Shannon diversity, the metabolite trajectory, the diet score or metabolite/diet-associated OTUs/taxa to predict MetS status (adjusting for age, sex, and family relatedness), and each component adjusting for age, BMI (except for BMI), sex, and family relatedness.

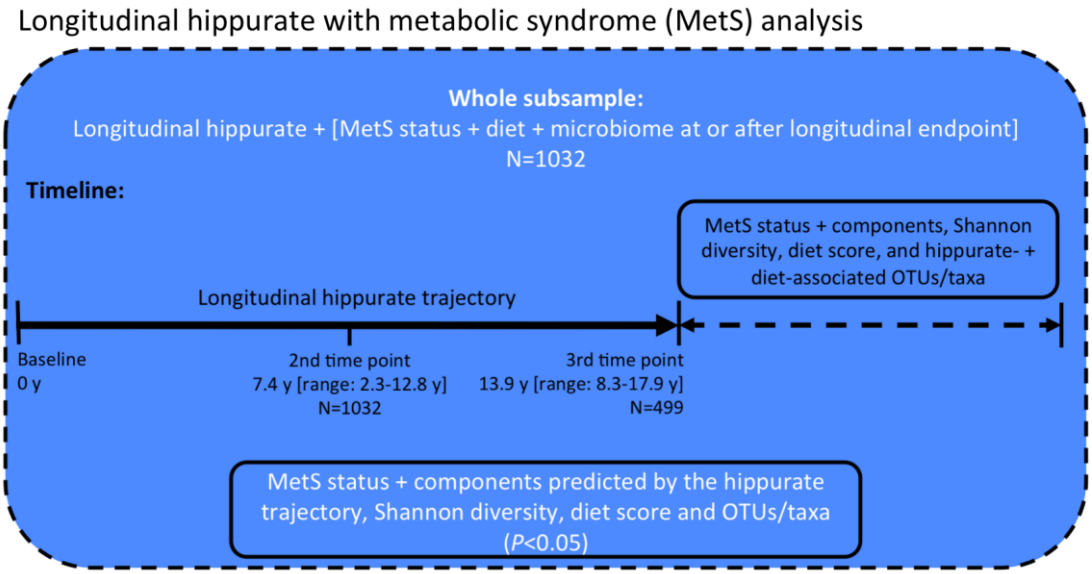
6.2.3.2.2 Metabolite trajectory association with MetS (and components) mediated by Shannon diversity, the diet score and specific OTUs/taxa

The aim of this analysis was to determine the degree to which the variance in MetS (and components) attributable to the metabolite trajectory was accounted for by the MetS (and components) association with Shannon diversity, the diet score or specific OTUs/taxa. The proportion of the variance of the MetS status and its components were determined for the metabolite trajectory after taking into account all covariates. This quantity is indicated as r^2_x. The proportion of the variance for the MetS status (and its components) explained by the metabolite trajectory was then calculated after taking into account the same covariates as above but also including, Shannon diversity, the diet score (if applicable) or associated OTUs/taxa (r^2_{xy}). The percentage of the longitudinal metabolite association mediated by Shannon diversity, the diet score and associated OTUS/taxa (r^2_y.) was calculated as the proportion of the variance of MetS status (and components) that is due to the longitudinal metabolite trajectory association with Shannon diversity, the diet score and associated taxa, namely $1 – (r^2_{xy}/r^2_x)$.

6.2.3.2.1 Confirmation of findings in twins discordant for Shannon diversity

Fifty-five MZ twin pairs discordant (≥ 1 SD) for Shannon diversity were used to confirm associations with diversity associations with the top metabolite cross-sectionally and longitudinally, the diet score and significant MetS components.

Figure 6-2. Flowchart of study design for part 2



6.3 Part 1 results

6.3.1 Microbiome diversity metabolomics associations

Eight metabolites significantly correlated with Shannon diversity in the discovery sample and five were validated in the validation sample (**Table 6-4**). Among those, hippurate, a benzoate metabolite, associated significantly (P<5.00x10^{-4}: 0.05/[5 metabolites x 20 food groups]) with intakes of fruit (0.012[0.002]; P=7.36x10^{-8}) and whole grains (0.013[0.003]; P=2.05x10^{-5}). Another benzoate metabolite 3-phenylpropionate was associated with fried foods (-0.045[0.009]; P=5.63x10^{-7}), whole grains (0.018[0.004]; P=2.71x10^{-6}) and fruit (0.010[0.002]; P=2.45x10^{-5}). Hippurate and 3-phenylpropionate were strongly correlated (r=0.51), although summing the two metabolites did not improve their association with Shannon diversity (hippurate R^2: 0.0258; 3-phenylpropionate R^2: 0.0122; and combined R^2: 0.0236), therefore for the remainder of the analysis I focused on hippurate.

6.3.2 Food intakes predict longitudinal hippurate trajectories

I next analysed longitudinal levels of hippurate (**Figure 6-3** shows the data for 25 randomly selected twins). Higher intakes of whole grains, coffee and fruit significantly (P<0.0025) predicted increasing hippurate trajectories (**Table 6-5**). All associations remained significant in a multivariate linear regression (**Table 6-5**) and together accounted for 5.3% of the variance in the hippurate trajectory. The nutrient profile of the score formed from these foods is shown in **Table 6-6**. I validated the hippurate diet score in the validation sample against hippurate (0.089[0.024]; P=3.21x10^{-4}) and Shannon diversity (0.040[0.019]; P=0.035), independently of hippurate. The diet score was moderately (h^2≥30%) heritable (A: 0.3782 [0.3024, 0.4485]; E: 0.6218 [0.5515, 0.6976]), with the AE model being the best fit.

Figure 6-3. Blood hippurate levels at 3 visits

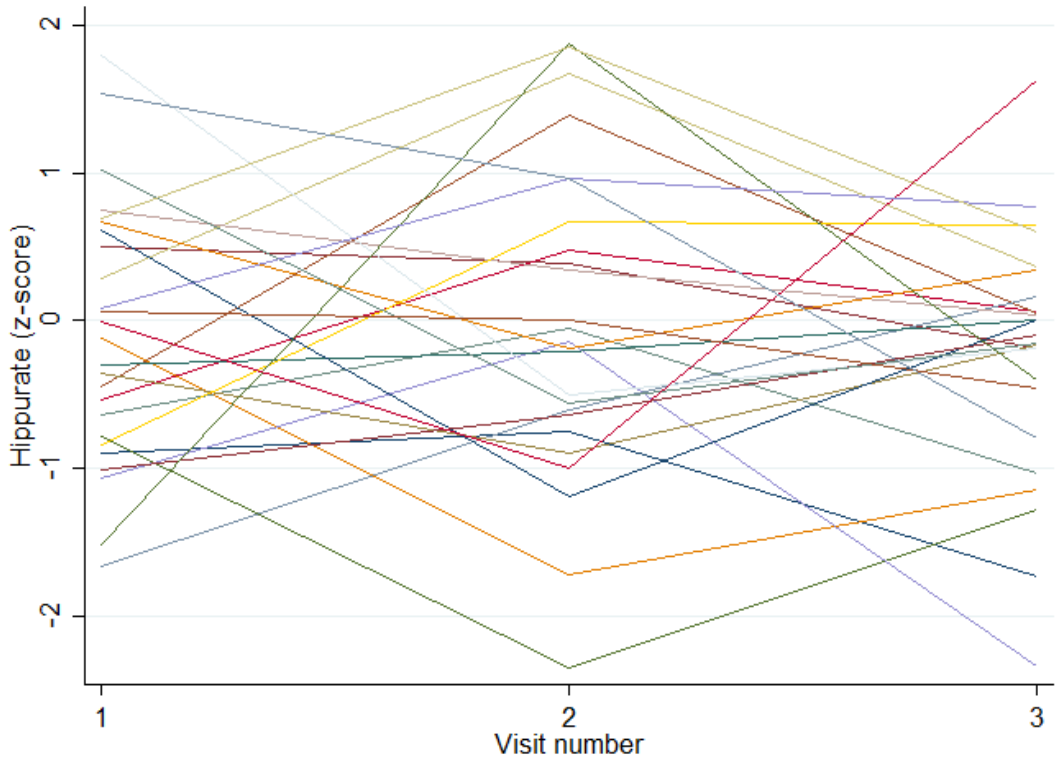


Figure 6-3 shows blood hippurate levels at 3 different visits for 25 randomly selected twins. The data were adjusted for age and BMI.

Table 6-4. Metabolites associated to Shannon diversity in the discovery sample, following backward stepwise linear regression and in the validation sample [(1)]

Metabolite	Super-pathway	Sub-pathway	Discovery (*n*=1529) beta (SE)	*P*	Validation (*n*=420) [(2)] beta (SE)	*P*
Hippurate	Xenobiotics	Benzoate metabolism	0.230 (0.040)	$3.72x10^{-8}$	0.238 (0.072)	0.001*
p-cresol sulfate	Amino acid	Phenylalanine & tyrosine metabolism	0.200 (0.040)	$9.90x10^{-8}$	0.179 (0.063)	0.005*
phenol sulfate	Amino acid	Phenylalanine & tyrosine metabolism	-0.200 (0.040)	$5.82x10^{-7}$	-0.121 (0.063)	0.055
Phenylacetylglutamine	Amino acid	Phenylalanine & tyrosine metabolism	0.180 (0.040)	$5.21x10^{-6}$	0.195 (0.062)	0.002*
3-phenylpropionate (hydrocinnamate)	Amino acid	Phenylalanine & tyrosine metabolism	0.160 (0.040)	$3.43x10^{-5}$	0.185 (0.084)	0.028*
4-ethylphenylsulfate	Xenobiotics	Benzoate metabolism	0.190 (0.050)	$5.12x10^{-5}$	0.062 (0.081)	0.441
Hyodeoxycholate	Lipid	Bile acid metabolism	-0.190 (0.050)	$8.66x10^{-5}$	-0.215 (0.089)	0.016*
Indolepropionate	Amino acid	Tryptophan metabolism	0.140 (0.040)	$9.20x10^{-5}$	0.093 (0.083)	0.262

*Statistically significant: *P*<0.05

(1) A linear regression was performed using Shannon diversity to predict levels of 292 metabolites adjusting for age, BMI, batch effects (and sex in the validation) and family relatedness.

(2) Statistically significant (*P*<$1.71x10^{-4}$) associations from the discovery group were validated in the validation group.

Table 6-5. Food intake and score predictions of longitudinal hippurate trajectories and following a backwards stepwise linear regression in a subsample of 788 twins

	Independent regression [(1)]		Multivariate regression [(2)]	
	Beta (SE)	*P*	Beta (SE)	*P*
Whole grains	$1.70x10^{-4}$ ($3.84x10^{-5}$)	$9.54x10^{-6}$	$1.58x10^{-4}$ ($3.83x10^{-5}$)	$3.90x10^{-5}$
Coffee	$1.03x10^{-4}$ ($2.82x10^{-5}$)	$2.73x10^{-4}$	$1.14x10^{-4}$ ($2.79x10^{-5}$)	$4.46x10^{-5}$
Fruit	$8.43x10^{-5}$ ($2.71x10^{-5}$)	$1.89x10^{-3}$	$7.90x10^{-5}$ ($2.70x10^{-5}$)	$3.48x10^{-3}$
Diet score	$1.07x10^{-3}$ ($1.71x10^{-4}$)	$7.18x10^{-10}$		

(1) Intakes of 20 food groups at baseline were used to predict the hippurate trajectory adjusted for age, BMI and sex. Statistical significance was defined as *P*<0.0025 (0.05/20 food groups). The diet score was formed by summing quartile ranked intakes of food groups significantly associated with the hippurate trajectory.

(2) Whole grain, coffee and fruit intake were included in a multivariate regression to predict the hippurate trajectory including age, BMI and sex as covariates.

Table 6-6. Nutrient profile and linear trends of the hippurate diet score according to tertiles of the hippurate diet score

	Tertile 1	Tertile 2	Tertile 3	Trend	
Nutrient	Mean (SD)	Mean (SD)	Mean (SD)	Beta (SE)	*P*
Energy (kcal)	2009.4 (562.8)	1847.1 (534.2)	1826.6 (498.7)	-90.77 (14.75)	$9.16x10^{-10}$
Fat (g/d)	72.2 (12.5)	67.2 (10.4)	62.5 (9.9)	-4.87 (0.3)	$9.54x10^{-55}$
Saturated FAs (g/d)	26.8 (6.6)	23.7 (5.8)	21.2 (5.2)	-2.76 (0.16)	$8.00x10^{-61}$
MUFAs (g/d)	24.2 (4.6)	22.1 (3.8)	20.2 (3.7)	-1.98 (0.11)	$2.19x10^{-64}$
PUFAs (g/d)	15.6 (5.2)	15.8 (4.3)	15.7 (4)	0.04 (0.13)	NS
trans-FAs (g/d)	1.8 (0.7)	1.5 (0.6)	1.3 (0.5)	-0.22 (0.02)	$2.47x10^{-37}$
Cholesterol (mg/d)	241.4 (92.2)	233.1 (78.5)	210.8 (70.7)	-15.34 (2.24)	$8.98x10^{-12}$
Protein (g/d)	79.3 (13)	80.5 (12.7)	80.4 (11.5)	0.56 (0.34)	NS
Carbohydrate (g/d)	221.6 (39.3)	232.2 (32)	245.2 (29.8)	11.83 (0.93)	$2.14x10^{-35}$
Starch (g/d)	112.3 (31.2)	108.8 (26.6)	111.2 (27)	-0.54 (0.78)	NS
Total sugars (g/d)	106.6 (30.6)	120.7 (28.2)	131.1 (26.2)	12.23 (0.78)	$8.44x10^{-52}$
Glucose (g/d)	17.8 (7.8)	23.4 (9.1)	27.1 (7.9)	4.64 (0.23)	$2.42x10^{-82}$
Fructose (g/d)	19.8 (9.1)	27.5 (11.2)	32.9 (9.8)	6.57 (0.28)	$1.02x10^{-107}$
Sucrose (g/d)	44.3 (18.9)	44.2 (14.2)	45.5 (12.8)	0.58 (0.43)	NS
Maltose (g/d)	3.4 (1.9)	3.1 (1.5)	3.3 (1.5)	-0.03 (0.05)	NS
Lactose (g/d)	18.2 (10.7)	17.7 (9.8)	17.3 (9.2)	-0.45 (0.27)	$9.79x10^{-2}$
NSP (g/d)	16.8 (5)	20.4 (4.9)	23.4 (5.2)	3.29 (0.14)	$9.35x10^{-110}$
Alcohol (g/d)	10.6 (14.8)	9.6 (11.9)	8.6 (10.5)	-0.95 (0.35)	$5.77x10^{-3}$
Water (g/d)	2395.3 (596.3)	2623.5 (575.6)	2817.4 (591.6)	210.9 (16.26)	$5.21x10^{-37}$
Sodium (mg/d)	2172.9 (487.6)	2263 (502.5)	2341.4 (468.7)	84.19 (13.47)	$4.98x10^{-10}$
Potassium (mg/d)	3526.7 (575.9)	3886 (576.2)	4174.1 (564.8)	323.4 (15.84)	$2.50x10^{-84}$
Chloride (mg/d)	3446.9 (750.3)	3603.9 (773)	3757.3 (712.8)	155.16 (20.65)	$8.56x10^{-14}$
Calcium (mg/d)	1015.3 (288)	1037.2 (276.3)	1037.9 (274.2)	11.17 (7.73)	NS
Magnesium (mg/d)	299.1 (48.6)	343.4 (48.3)	377.1 (50.4)	38.93 (1.36)	$2.03x10^{-151}$
Phosphorous (mg/d)	1388.4 (223.7)	1476.2 (213.8)	1528.9 (207.9)	70.11 (5.96)	$5.50x10^{-31}$
Iron (mg/d)	11.5 (2.5)	12.8 (2.7)	13.7 (3)	1.08 (0.08)	$1.64x10^{-43}$
Copper (mg/d)	1.5 (0.5)	1.6 (0.5)	1.6 (0.4)	0.08 (0.01)	$5.57x10^{-9}$
Zinc (mg/d)	9.8 (1.7)	10.2 (1.6)	10.3 (1.5)	0.27 (0.04)	$2.00x10^{-9}$
Manganese (mg/d)	3.7 (1.1)	4.1 (1.1)	4.5 (1.1)	0.41 (0.03)	$2.02x10^{-39}$
Iodine (ug/d)	199.3 (74.8)	205.7 (72.6)	206.1 (67.2)	3.35 (1.98)	$9.12x10^{-2}$
Retinol (ug/d)	610 (776.9)	500.6 (508.6)	439.2 (425.7)	-85.14 (16.18)	$1.58x10^{-7}$
Carotene (ug/d)	4987.8 (3370.1)	5754.6 (3645.7)	5896.5 (3627.5)	451.48 (98.45)	$4.80x10^{-6}$
Vitamin D (ug/d)	2.3 (1)	2.5 (1)	2.5 (1.2)	0.1 (0.03)	$5.47x10^{-4}$
Vitamin E (mg/d)	10.1 (3.2)	11.2 (3.2)	11.7 (3.1)	0.78 (0.09)	$9.48x10^{-19}$
Thiamin (mg/d)	1.6 (0.4)	1.7 (0.4)	1.8 (0.4)	0.1 (0.01)	$1.15x10^{-21}$
Riboflavin (mg/d)	2.2 (0.7)	2.2 (0.7)	2.3 (0.6)	0.02 (0.02)	NS
Niacin (mg/d)	19.8 (4.9)	21.2 (4.7)	23 (4.6)	1.6 (0.13)	$2.67x10^{-33}$
Tryptophan (mg/d)	16.6 (2.7)	16.8 (2.6)	16.9 (2.3)	0.17 (0.07)	$1.59x10^{-2}$
Vitamin B6 (mg/d)	2.3 (0.6)	2.5 (0.6)	2.7 (0.5)	0.16 (0.02)	$6.41x10^{-26}$
Vitamin B12 (mg/d)	6.2 (3)	6.1 (2.4)	5.9 (2.2)	-0.16 (0.07)	$1.87x10^{-2}$
Folate (ug/d)	358 (106.9)	388.7 (113.4)	416 (110.1)	28.97 (3.05)	$6.00x10^{-21}$
Pantothenate (mg/d)	5.7 (1.6)	6.2 (5.3)	6.2 (1.8)	0.24 (0.09)	$1.11x10^{-2}$
Biotin (ug/d)	40.9 (9.8)	45.4 (9.1)	49.4 (8.9)	4.25 (0.26)	$4.40x10^{-58}$
Vitamin C (mg/d)	126.8 (58.5)	169.6 (72.2)	200.6 (74.1)	36.88 (1.91)	$1.37x10^{-76}$

NS= not significant: P>0.05. FA: fatty acid, MUFAS: monounsaturated fatty acids, PUFAS: polyunsaturated fatty acids, NSP: non-starch polysaccharides

Tertile 1: score 0-3; tertile 2: score 4-5; tertile 3: scores 6-9. Linear trend determined by using the tertile of the hippurate diet score as a predictor of the energy-adjusted nutrient intake.

6.3.3 Utility of hippurate as a marker of diversity and the hippurate diet score as a marker of hippurate

Overall hippurate measured at a single time point correctly classified 64.64% of individuals into the upper and lower tertiles of Shannon diversity (sensitivity: 68.57%; specificity: 60.71%). The AUC was 0.682 (0.620, 0.745) (**Figure 6-4**). The hippurate diet score correctly classified 60.93% of individuals into the upper and lower tertiles of hippurate levels (sensitivity: 64.03%; specificity: 57.86%). The AUC was 0.649 (0.585, 0.713) (**Figure 6-5**). This suggests both models had low accuracy.

Figure 6-4. Receiver operating characteristic curve for the ability of hippurate to predict the upper and lower tertiles of Shannon diversity in the validation sample

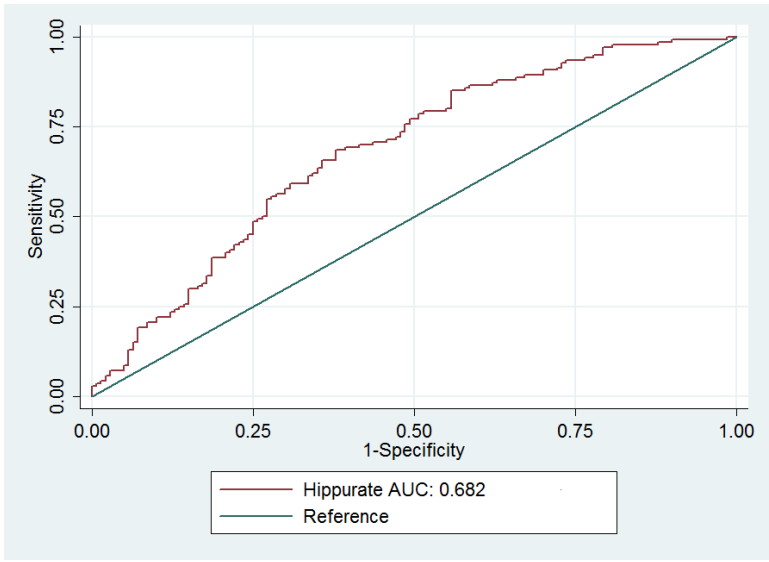
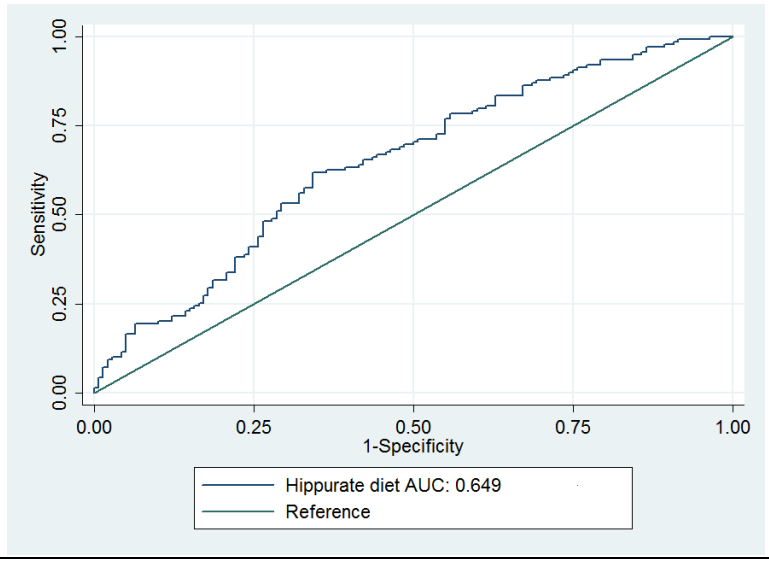


Figure 6-5. Receiver operating characteristic curve for the ability of the hippurate diet score to predict the upper and lower tertiles of blood hippurate in the validation sample



6.3.4 Associations with hippurate and the hippurate diet score across diversity metrics

I confirmed the same directional effects for other diversity metrics for hippurate and the diet score (**Table 6-7**).

Table 6-7. Associations between hippurate and the hippurate diet score across diversity metrics in the whole sample

	Hippurate [2]		Hippurate diet score [3]	
Diversity metric [1]	Beta(SE)	*P*	Beta(SE)	*P*
Observed species	0.158 (0.024)	$9.55x10^{-11}$	0.086 (0.023)	$2.17x10^{-4}$
Shannon	0.160 (0.025)	$2.16x10^{-10}$	0.108 (0.023)	$3.03x10^{-6}$
Simpson	0.082 (0.026)	0.002	0.062 (0.021)	0.003
Chao1	0.060 (0.023)	0.011	0.023 (0.023)	*NS*

NS= not significant: P>0.05.

(1) Standardized to have mean 0, SD 1.
(2) Hippurate associations with diversity metrics adjusted for sex, age, BMI, metabolite batch and family relatedness.
(3) Hippurate diet score associations with diversity adjusted for sex, hippurate, age, BMI, metabolite batch and family relatedness.

6.3.5 OTU and collapsed taxa associations with hippurate

Thirty OTUs and sixteen collapsed taxa (**Figure 6-6; Appendix E Tables 1 and 2**) were significantly associated with blood levels of hippurate ($P<8.61x10^{-5}$ [OTUs]-$1.47x10^{-3}$ [phylum]). At the phylum level, the Firmicutes were associated with reduced levels of hippurate in blood, suggesting reduced diversity. Within the Firmicutes phylum, 28 OTU associations were assigned to the order Clostridiales, though not all associations were in the same direction. Within the order Clostridiales, higher hippurate levels were associated with increased abundances of two OTUs assigned to the family Clostridiaceae and 10 OTUs assigned to the family Ruminococcacea (including 2 *Oscillospira* genus OTUs, 1 *Ruminococcus* genus OTU and 1 *Faecalibacterium prausnitzii* OTU).

Increased abundances of 14 OTUs assigned to the family Lachnospiraceae were associated with reduced hippurate levels. Associations within the Lachnospiraceae family remained strong and in the same direction within the collapsed taxonomies, including at the family level. In particular, within the Lachnospiraceae family there were associations with OTUs assigned to the genera *Ruminococcus* (2 OTUs), *Blautia* (4 OTUs), *Dorea* (2 OTUs) and *Roseburia* (1 OTU), these OTUs were likely driving associations within their respective collapsed genera.

At the class level, lower hippurate levels were associated with increased abundance of Erysipelotrichi, a trend that continued at the order, family and genus levels with

Erysipelotrichales, Erysipelotrichaceae, and *Eubacterium*, respectively. Interestingly there were no OTUs associated to blood hippurate that were assigned to this class.

Outside of the Firmicutes phyla, increased abundances of one OTU assigned to the genus *Actinomyces* was associated with increased blood hippurate levels, this association was likely driving associations within the family Actinomycetaceae and order Actinomycetales collapsed taxa. Increased levels of hippurate were also associated with increased abundances of an OTU belonging to the family Rikenellaceae of the Bacteroidetes phyla and reduced abundances of the collapsed genus *Ralstonia* of the Proteobacteria phyla.

Following a backward stepwise linear regression using $P<0.05$ as the cut-off threshold, together the remaining OTUs accounted for 58.0% of the variance in Shannon diversity and 7.1% of the variance in hippurate levels (adjusted for diversity).

6.3.6 OTU and taxa associated with both hippurate and the hippurate diet score

Five OTUs and five taxa were associated with hippurate were also associated with the diet score in the same direction (**Table 6-8**). Specifically, reduced abundances of one OTU assigned to the genus *Actinomyces* was associated with increased scores on the diet, similar to hippurate this trend was significant both the family and order levels. At the genus level, reduced abundances of *Ruminococcus* (including 2 OTUs) and *Eubacterium* were associated with increasing diet scores. Moreover, increased diet scores were associated to increased abundances of OTUs assigned to the species *Faecalibacterium prausnitzii* and the genus *Clostridiales*. Most associations appeared to be primarily driven by intakes of fruit and whole grains. Though higher abundances of one Clostridiales OTU were associated with increased coffee intake.

I did not identify any significant interactions between the diet score and hippurate/diet-associated microbes independently, nor when abundances were summed, in the prediction of blood hippurate levels.

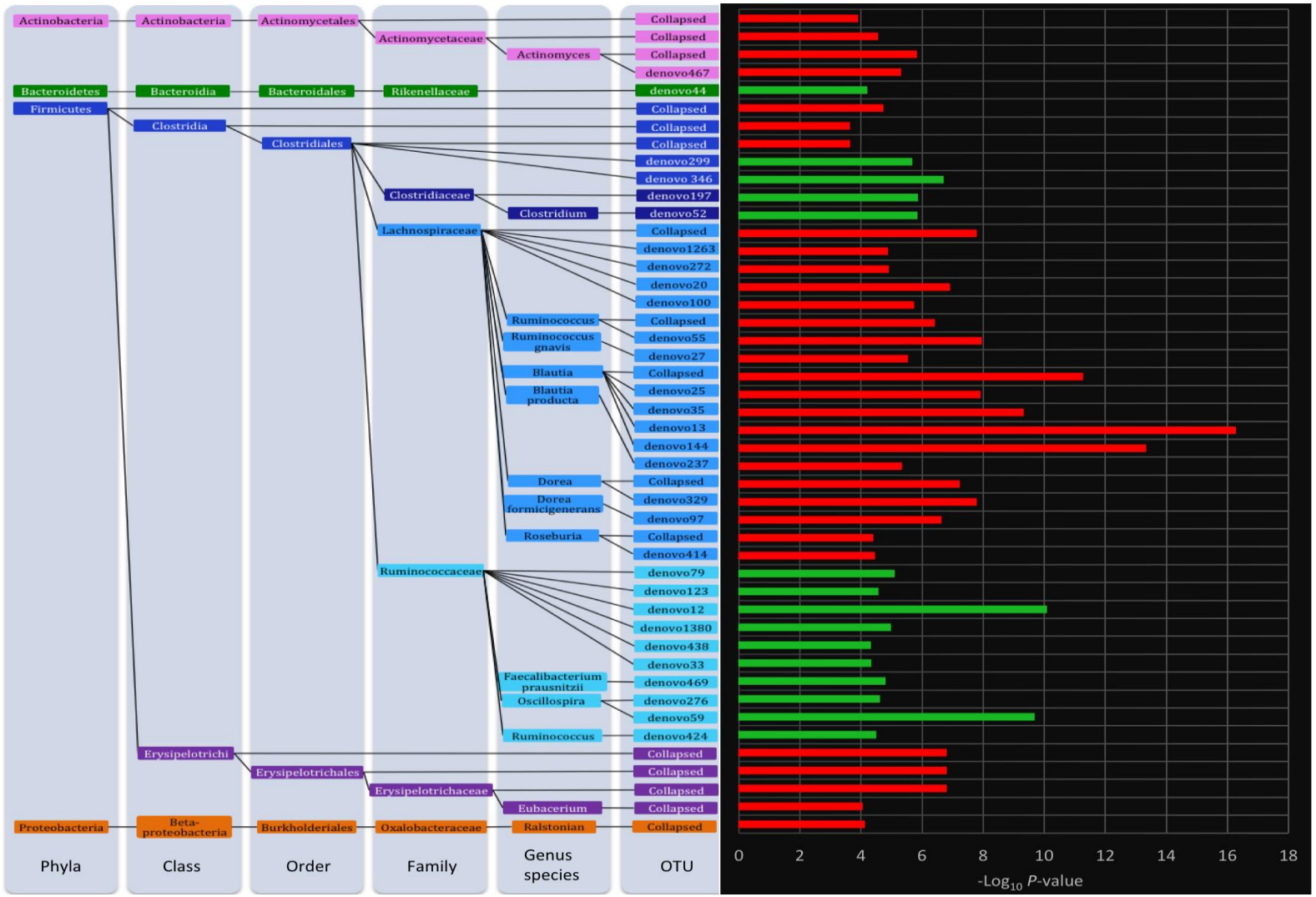


Figure 6-6. OTU and collapsed taxonomic associations with hippurate
Associations between blood hippurate and microbiome variables are represented by the histogram bars on the right side of the plot. The histogram bars represent the $-\log_{10}$ of the *P*-value of the regression and the colour of the bars indicates the direction of association: *green*, positive; *red*, negative.

Table 6-8. List of taxa associated with hippurate, the hippurate diet score and foods [1]

				Hippurate		Diet score		
Order	Family	Genus species	OTU ID [2]	Beta (SE)	*P*	Beta (SE)	*P*	Foods [3] *P*<0.05
Actinomycetales			Collapsed	-0.083(0.022)	$1.31x10^{-4}$	-0.035(0.011)	$1.67x10^{-3}$	Fruit: -0.004(0.002) WG: -0.007(0.003)
Actinomycetales	Actinomycetaceae		Collapsed	-0.089(0.021)	$2.89x10^{-5}$	-0.036(0.011)	$1.70x10^{-3}$	Fruit: -0.004(0.002) WG: -0.007(0.003)
Actinomycetales	Actinomycetaceae	*Actinomyces*	Collapsed	-0.101(0.021)	$1.55x10^{-6}$	-0.045(0.011)	$5.71x10^{-5}$	Fruit: -0.005(0.002) WG: -0.008(0.003)
Actinomycetales	Actinomycetaceae	*Actinomyces*	denovo467	-0.099(0.022)	$5.14x10^{-6}$	-0.051(0.011)	$2.81x10^{-6}$	Fruit: -0.005(0.002) WG: -0.009(0.003)
Clostridiales			denovo299	0.113(0.024)	$2.21x10^{-6}$	0.044(0.010)	$9.76x10^{-6}$	Coffee: 0.013(0.002)*
Clostridiales	Lachnospiraceae	*Ruminococcus*	Collapsed	-0.111(0.022)	$4.03x10^{-7}$	-0.038(0.011)	$6.35x10^{-4}$	Fruit: -0.005(0.002) WG: -0.008(0.003)
Clostridiales	Lachnospiraceae	*Ruminococcus*	denovo55	-0.123(0.021)	$1.17x10^{-8}$	-0.054(0.011)	$2.79x10^{-6}$	Fruit: -0.006(0.002)* WG: -0.009(0.003)
Clostridiales	Lachnospiraceae	*Ruminococcus gnavis*	denovo27	-0.107(0.023)	$3.04x10^{-6}$	-0.064(0.011)	$1.99x10^{-8}$	Fruit: -0.006(0.002)* WG: -0.009(0.003)
Clostridiales	Ruminococcaceae	*Faecalibacterium prausnitzii*	denovo469	0.100(0.023)	$1.66x10^{-5}$	0.034(0.010)	$9.24x10^{-4}$	WG: 0.007(0.003)
Erysipelotrichales	Erysipelotrichaceae	*Eubacterium*	Collapsed	-0.083(0.021)	$9.30x10^{-5}$	-0.040(0.012)	$6.12x10^{-4}$	Fruit: -0.004(0.002) WG: -0.010(0.003)*

*= statistically significant: *P*<0.0017; WG: whole grain products

(1) Microbiome OTUs and collapsed taxa significantly associated with both hippurate and the hippurate diet score are shown. Associations were adjusted for covariates (age, Shannon Index, metabolite batch, BMI, sex and family relatedness) and multiple testing using Bonferroni correction. Hippurate diet score associations were also adjusted for hippurate.

(2) OTU ID assignment is specific to the TwinsUK cohort.

(3) All foods included in the hippurate diet score were fitted into a backwards stepwise linear regression using *P*<0.05 as the cut-off threshold with each taxa associated to both hippurate and the diet score. Results displayed are the betas with standard errors of foods at least nominally associated (*P*<0.05). Statistical significance was defined as *P*<0.0017 (Bonferronil: 0.05/[10 taxa x 3 foods]).

6.4 Part 2 results

6.4.1 Relationship of diversity, the hippurate trajectory and diet to MetS and its components

Longitudinal hippurate trajectories were significantly associated with Shannon diversity, independently of diet and covariates in a subsample of 1032 twins (15.736[1.96]; P=4.95x10^{-15}), moreover the hippurate trajectory accounted for 6.5% of the variance in Shannon diversity.

Figure 6-7 shows the results of the analysis for associations between diversity, the hippurate trajectory and MetS (**Appendix E Tables 3 and 4** contain the full results). Higher Shannon diversity (R^2: 0.0105) and an increasing hippurate trajectory (R^2: 0.0054) were associated with a reduced risk of having MetS (**Figure 6-7a**). The variance in MetS attributable to the hippurate trajectory was accounted for 61.1% by Shannon diversity. The variance in MetS attributable to the hippurate trajectory was accounted for from 3.7% (*Eubacterium* genus) to 51.9% (Clostridiales OTU) for the five collapsed taxa and 3 of the OTUs that were associated to both hippurate and the diet (**Figure 6-7b**).

Increased Shannon diversity and the hippurate trajectory were significantly associated with a reduced BMI (diversity R^2: 0.0288) and TG (diversity R^2: 0.0213; **Figure 6-7a**). Increased Shannon diversity was also associated with higher HDL cholesterol (R^2: 0.0126). Shannon diversity accounted for 57.6% of the variance in BMI (R^2: 0.0288) attributable to the hippurate trajectory, the impact of specific OTUs/taxa ranged from 22.6% (*Ruminococcus* genus) to 35.4% (*Faecalibacterium prausnitzii*). Shannon diversity accounted for 63.2% of the variance in TG (R^2: 0.0087) attributable to the hippurate trajectory, the impact of specific OTUs/taxa ranged from 3.4% (*Eubacterium* genus) to 48.3% (Clostridiales OTU).

Figure 6-7. Associations between diversity, the hippurate trajectory, diet and OTUs and collapsed taxa with MetS status and its components

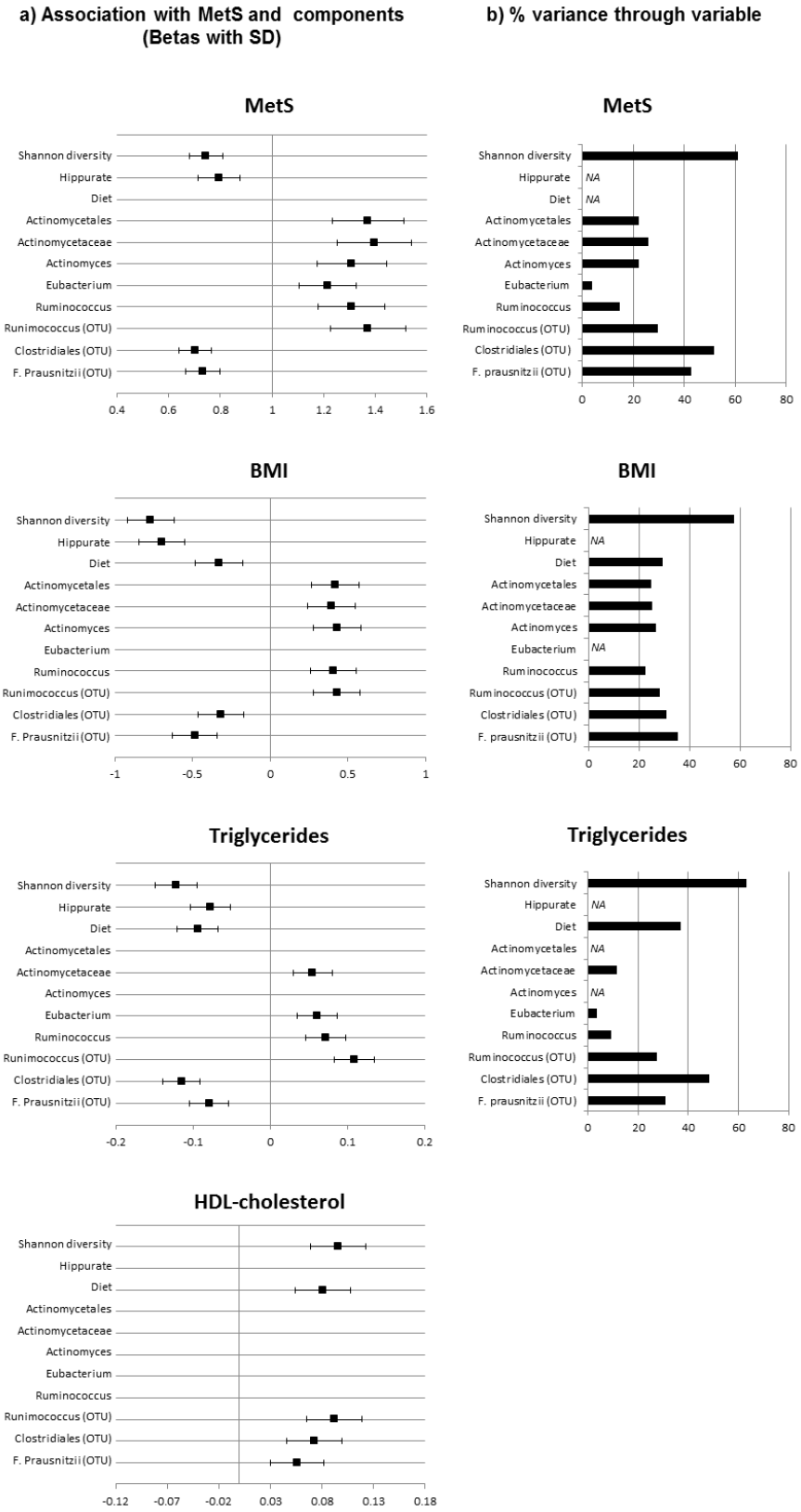


Figure 6-7a shows the associations between MetS, BMI, triglycerides and HDL-cholesterol with Shannon diversity, the hippurate trajectory, the diet score and taxa significantly associated with both hippurate, the diet score and MetS represented as betas with SEs; all variables have been standardized. Figure 6-6b shows the percentage variance the MetS, BMI and triglycerides are attributable to the hippurate trajectory accounted for through each applicable variable.
Abbreviations: MetS, metabolic syndrome; HDL, high density lipoprotein; BMI, body mass index.

6.4.2 Confirmation of results in discordant twins

I identified 55 MZ twin pairs who were discordant (≥1 SD apart) for Shannon diversity. Associations between Shannon diversity and hippurate cross-sectionally and the longitudinal trajectory were significant (P<0.05), associations with all other variables were in the same direction as in the whole group analysis (**Table 6-9**), except for HDL-cholesterol.

Table 6-9. Associations between hippurate (discovery), the hippurate trajectory, MetS status and components in MZ twins discordant for diversity

Variable	Beta(SE)	*P*	R^2
Hippurate (discovery)	0.208 (0.081)	0.013	0.0607
Hippurate trajectory	0.478 (0.078)	$9.53x10^{-8}$	0.1768
Hippurate diet score	0.149 (0.099)	0.137	0.0136
MetS status*	1.046 (0.299)	0.875	0.0004
BMI	0.021 (0.063)	0.737	0.0010
HDL-cholesterol	-0.035 (0.069)	0.612	0.0022
TG	-0.073 (0.060)	0.233	0.0118

MetS, metabolic syndrome; HDL, high density lipoprotein; TG, triglycerides
A linear regression was conducted using diversity to predict hippurate (discovery), the hippurate trajectory, and MetS status and components in the MZ discordant (1 SD apart in diversity) twin sample.
*Statistical results show the odds ratio. Variables were standardized to have mean=0, SD=1.

6.5 Discussion

In summary circulating levels of hippurate were significantly and positively associated with gut microbiome diversity in all twins and confirmed in MZ twins discordant for diversity. Blood hippurate levels longitudinally and cross-sectionally were also associated with food intakes, specifically self-reported intakes of fruit and whole grains. Longitudinal changes in blood hippurate (reflecting in part food polyphenol content) were also strongly associated with Shannon diversity, accounting for 17.7% of the variance in diversity in MZ discordant twins. Blood hippurate levels and reported intakes of associated foods were associated with 5 OTUs and 5 collapsed taxa independent of diversity. Both an increasing Shannon diversity and higher hippurate trajectory were associated with a lower risk of MetS, which was primarily accounted for by their relationship to BMI and TG, moreover the 5 collapsed taxa and 3 OTUs were associated with MetS.

Metabolite associations with diversity

I identified and validated in an independent sample of male and female twins 5 blood metabolites associated with gut microbiome diversity: hippurate, p-cresol sulfate, phenylacetylglutamine and 3-phenylpropionate which were associated positively; and negatively with hyodeoxycholate. All of these metabolites are derived from or modified by bacterial metabolism.

P-cresol sulfate (also p-cresyl sulfate) and phenylacetylglutamine are potentially toxic uremic solutes formed from the putrefication by colonic bacteria of dietary proteins that have escaped digestion in the small intestine. Our group previously showed p-cresol sulfate and phenylacetylglutamine to be strongly associated with reduced estimated glomerular filtration rate, moreover 52 OTUs assigned to the genera *Christensenellaceae*, *Ruminococcaceae* and *Lachnospiraceae* were associated with phenylacetylglutamine and 3 OTUs assigned to the genera Ruminococcaceae and Lachnospiraceae with P-cresol sulfate (Barrios et al., 2015). Hyodeoxycholate is a secondary bile acid, produced from intestinal bacterial metabolism. In healthy individuals hyodeoxycholate is metabolised by glucuronidation in the human liver and kidneys, a pathway for toxin elimination (Perreault et al., 2013). In individuals with cholestatic liver disease or intestinal malabsorption increased levels of hyodeoxycholate are excreted in the

urine (Sacquet et al., 1983), suggesting hyodeoxycholate may act as a marker of poor liver function. Which is an interesting observation considering the relationship between some metabolic diseases and poor liver function.

Increased blood 3-phenylpropionate and hippurate were associated with increased diversity and also intakes of fruits and whole grains. Both 3-phenylpropionate and hippurate are derived from gut microbial metabolism of polyphenols to benzoates and are significantly correlated (r=0.51), though hippurate accounted for a greater degree of variance in diversity therefore I decided to focus on it. Hippurate is emerging as a key mammalian-microbial co-metabolite. It is a glycine conjugate of benzoic acid formed in the mitochondria of the liver (Gatley and Sherratt, 1977) and kidneys (Temellini et al., 1993), through gut bacterial production of benzoic acid from dietary components, primarily polyphenols (Gonthier et al., 2003). Supporting my findings, human feeding studies of foods with a high polyphenol content, such as teas, fruit and coffee, have shown to increase urinary output of hippurate (Gonthier et al., 2003, Walsh et al., 2007). It should be noted however that the ROC analysis showed hippurate was a poor marker of Shannon diversity, therefore future studies should aim to identify biomarkers with better predictive performance.

Microbiome associations with hippurate but not associated with diet

To date, the microbiome profile associated with blood hippurate levels has not been well characterised. Within the TwinsUK dataset I found overall higher hippurate levels were associated to both increased and reduced abundances of 30 OTUs and 16 collapsed taxa, independently of Shannon diversity. At the phylum level, increased levels of hippurate were associated with reduced abundances of Firmicutes. The Firmicutes, which primarily consist of gram-positive bacteria, are one of the dominating phyla of the adult microbiome. One of the first large-scale microbiome studies suggested that a high Firmicutes to Bacteroidetes ratio is characteristic of obesity (Ley et al., 2006), although I did not investigate this ratio. A handful of studies have found reduced Firmicutes levels with increasing BMI (Duncan et al., 2008, Tims et al., 2013, Escobar et al., 2014). Moreover, increased faecal SCFAs have been shown in obese compared to lean subjects which also correlated with a higher ratio of Firmicutes to Bacteroides/Prevotella (Fernandes et al., 2014). Interestingly, polyphenol (concord grape extract) feeding to mice has recently shown to prevent obesity induced by a high fat diet in part by decreasing the Firmicutes/Bacterodetes ratio (Roopchand et al., 2015). Although I did not

find a relationship between total Firmicutes and dietary factors, it is possible our dataset was not powerful enough to detect associations.

Increased hippurate levels were associated to both increased and decreased abundances of bacteria within lower taxonomies of the Firmicutes phyla. Twenty-eight associations with hippurate were with OTUs/taxa belonging to the Clostridiales order which was negatively associated overall. The Clostridiales order belonging to the class Clostridia, are obligate anaerobes found primarily in soil. Directions within lower taxonomic levels of Clostridiales were not consistent with the collapsed order. For instance, increased abundances of two OTUs belonging to the Clostridiales order and 2 OTUs within the Clostridiaceae family were associated with increased blood hippurate levels. OTUs within the Clostridiales order were related to diet and MetS and will be discussed in more detail below.

Notably, the direction of associations between hippurate and OTUs/taxa within the Ruminococcaceae (10 associations) and Lachnospiraceae (14 associations) families of the order Clostridiales, were consistently increased and decreased, respectively. Both Lachnospiraceae and Ruminococcaceae are involved in butyrate production (Vital et al., 2014). In a previous study conducted on the TwinsUK dataset, higher abundances of multiple OTUs assigned to Lachnospiraceae were associated with increased eGFR and reduced levels of blood phenylacetylglutamine (Barrios et al., 2015), a metabolite that was also related to diversity. Another study found that higher abundances of Ruminococcaceae and Lachnospiraceae were associated with increased levels of urinary 3-indoxyl-sulfate and better outcomes in patients at risk of developing gastrointestinal (GI) graft-versus-host-disease (Weber et al., 2015). Moreover, lower abundances of both families are associated with inflammatory bowel disease (Kostic et al., 2014). I did not identify associations with hippurate or other polyphenol-related metabolites in the literature. It is interesting and perhaps noteworthy that I identified clear opposing directions between hippurate levels and these families.

Higher levels of hippurate were associated with reduced abundances of the class Erysipelotrichi, and associations were consistent at the order, family and genus levels with Erysipelotrichales, Erysipelotrichaceae, and *Eubacterium*, respectively. Elevated levels of *Clostridium ramosum*, a member of Erysipelotrichi, have shown in mice fed a high-fat, obesogenic diet, which the authors suspected was due to increasing nutrient absorption (Woting et al., 2014). Moreover, on a 2-month inpatient study of 15 females fed a choline-deficient diet it

was found that faecal levels of of Erysipelotrichi and Gammaproteobacteria were related to changes in liver fat (Spencer et al., 2011).

Outside of the Firmicutes phyla, lower abundances of Actinobacteria were observed with increasing hippurate, these bacteria were further found to be associated with diet and MetS as discussed below. These results provide a microbiome profile associated with hippurate levels although, the complex metabolic relationships require further investigation. As I found diet to be significantly associated with food intakes, I investigated the relationship between these hippurate-associated OTUs and collapsed taxa.

Microbiome associations with hippurate and the hippurate diet

Three of the OTUs and 5 collapsed taxonomies associated to both hippurate and the hippurate diet (independently of diversity), moreover the variance in MetS attributable to the hippurate trajectory that was accounted for by these OTUs and taxa was between 3.7% and 51.9% (**Table 6-10**). These included the increased abundances of the Actinomycetaceae family, *Eubacterium* and *Ruminococcus* being associated with an increased risk (and reduced blood hippurate and diet score) and OTUs within the order Clostridiales and of the species *Faecalibacterium prausnitzii* associated with a reduced MetS risk (and increased blood hippurate and diet score). A detailed discussion of each observation is given below.

Table 6-10. Direction of associations between OTUs/collapsed taxa and hippurate, food intakes, MetS and its components.

Microbiome	Hippurate	Fruit	WG	Coffee	MetS	BMI	TG	HDL
Actinomycetales	↓	↓	↓		↑	↑		
Actinomycetaceae	↓	↓	↓		↑	↑	↑	
Actinomyces	↓	↓	↓		↑	↑		
Eubacterium	↓		↓		↑		↑	
Ruminococcus	↓	↓	↓		↑	↑	↑	
***Ruminococcus* (OTU)**	↓	↓	↓		↑	↑	↑	↑
Clostridiales (OTU)	↑			↑	↓	↓	↓	↑
***Faecalibacterium prausnitzii* (OTU)**	↑		↑		↓	↓	↓	↑

Abbreviations: WG, whole grain; MetS, metabolic syndrome; BMI, body mass index; TG, triglycerides; HDL, high density lipoprotein cholesterol.

An increasing hippurate trajectory associated with a lower MetS risk

Overall, increased Shannon diversity and the hippurate trajectory were associated with a reduced risk of having MetS. Supporting this, previous studies have shown reduced gut bacterial diversity has been found in metabolic conditions, including atherosclerosis (Koren et

al., 2011), obesity (Turnbaugh et al., 2009) and with phenotypes characteristic of the metabolic syndrome (Le Chatelier et al., 2013). Interestingly, most of the variance in MetS that was attributable to the hippurate trajectory was accounted for by Shannon diversity (61.1%). An increasing hippurate trajectory (accounting for BMI) was similarly associated with a lower BMI and TG (variance 57.6% and 63.2% by Shannon diversity, respectively) at endpoint.

Previous studies that examined urinary excretion or serum levels of hippurate have shown reduced urinary hippuric acid excretion in obesity (Shearer et al., 2008, Waldram et al., 2009, Calvani et al., 2010), though mainly in animal models. To my knowledge, no human studies to date have investigated specifically the relationship between the MetS and markers of dietary polyphenol consumption, such as hippurate. Though the diet score was not associated with MetS, it was associated with a lower BMI for which the hippurate trajectory (60.4%) and Shannon diversity (45.3%) contributed particularly strongly (**Appendix E Table 5**), suggesting that these factors were important mediators between of the effect of diet on obesity and consequential metabolic risk.

Increased abundances of the Actinomycetaceae family, and Eubacterium and Ruminococcus genera associate with MetS risk

The Actinomycetaceae family are typical commensals within the oral cavity. Elevated levels of Actinomycetaceae within the gut have been implicated in conditions where stomach acid production has been compromised, such as with proton pump inhibitor use (Imhann et al., 2016). The metabolic implications of increased abundance of Actinomycetaceae in the gut are not clear at this time, though higher levels within the oral cavity contribute to periodontitis and correlate with reduced insulin sensitivity (Demmer et al., 2015). In rare cases an *Actinomyces* overgrowth contribute to an infection within the gut (abdominal actinomycosis) through forming filamentous branches that grow through damaged mucosal tissue penetrating the gut barrier, forming abscesses and fistula (Bonnefond et al., 2016). As mucosal inflammation and bacterial translocation are emerging as important factors in the aggravation of MetS, *Actinomyces* should be explored further. The relationship between Actinomycetaceae and the foods forming the hippurate score is not entirely clear, though these findings suggest a healthy diet may prevent the proliferation of Actinomycetaceae.

Increased abundance of the genus *Eubacterium* was significantly associated with both reduced hippurate and the diet score (in particular whole grain intake), and higher risk of MetS.

The *Eubacterium* genus is within the family Eubacteriaceae of the Firmicutes phyla. Effects at the *Eubacterium* genus level have not been clarified, though different species belonging to the same genus may have different metabolic capacities and therefore be enriched by varying dietary exposures. Contrary to my findings others have shown feeding of whole grains (Martinez et al., 2013) and switching from a Western to plant-based diet (David et al., 2014) have shown to enrich abundances of *Eubacterium rectale*. More studies are needed to outline effects at the *Eubacterium* genus level or with other species.

Higher abundances of the genus *Ruminococcus* (one OTU and collapsed taxonomy) were associated with MetS risk and in particular higher levels of both HDL and TG. *Ruminococcus* are most abundant in one of the 3 major, though controversial (Jeffery et al., 2012) enterotypes (Arumugam et al., 2011). Functional analyses have shown the *Ruminococcus* enterotype to be enriched for genes that produce membrane sugar transporters (Arumugam et al., 2011), which may predispose to weight gain. *Ruminococcus* are also able to ferment complex carbohydrates (Walker et al., 2011) and produce alcohol (Christopherson et al., 2014). Previous studies have not found the *Ruminococcus* enterotype to be consistently modified by diet (Wu et al., 2011a, Claesson et al., 2012). I found increased *Ruminococcus* abundances were shown to be associated with a lower fruit intake. Following 12 weeks feeding of *schisandra chinensis* fruit, high in flavonoids, to obese women was previously been shown to reduce *Ruminococcus* abundances (Song et al., 2015). Increased abundances of *Ruminococcus* were also associated significantly with increased severity of NAFLD lesions (Boursier et al., 2016). Supporting the strong association between TG and *Ruminococcus*, uncultured phylotypes of the *Ruminococcus gnavus*-group correlated positively with polyunsaturated serum TGs of dietary origin (Lahti et al., 2013). The *Ruminococcus* genus contains species with differential effects on metabolism, moreover dietary effects are likely quite complex, for instance increased consumption of foods of animal origin increase abundances of *Ruminococcus gnavus*, but decrease *Ruminococcus bromii* and *Ruminococcus callidus* (David et al., 2014).

Increased abundances of OTUs belonging to the order Clostridiales and Faecalibacterium prausnitzii associate with protection from MetS

Increased abundances of OTUs within the order Clostridiales and species *Faecalibacterium prausnitzii* were associated with hippurate, dietary components and reduced risk of MetS. The

Clostridiales OTU was particularly strongly associated to coffee intake and TG. The gut microbiota play an important role in the metabolism of chlorogenic acids found in high concentrations in coffee. Significantly elevated levels of the *Clostridium coccoides-Eubacterium rectale* group have been shown following the incubation of human faecal microbiota with coffee samples (Mills et al., 2015). Due to the high polyphenol content coffee intake has been thought to have beneficial effects on metabolic health. In a large observational study including 93,179 individuals high coffee intake was associated with a low risk of obesity, MetS and type 2 diabetes, though this was not supported by Mendellian Randomization using five genetic variants associated with coffee intake (Nordestgaard et al., 2015).

Increased abundances of the *Faecalibacterium prausnitzii* OTU were strongly inversely associated with BMI and mildly with higher whole grain intake. *Faecalibacterium prausnitzii* is an important gut commensal accounting for 5% of the total faecal microbiota (Arumugam et al., 2011) that ferments dietary fibre to short-chain fatty acids (including butyrate). In our sample *Faecalibacterium prausnitzii* was strongly associated with Shannon diversity (Beta[SE]: 0.250[0.015]; P=5.09x10^{-59}) and most strongly related to BMI of all microbes related to hippurate and the diet. This clear effect reflects previous findings where *Faecalibacterium prausnitzii* was shown to correlate with microbiome gene count and predict weight loss over time (Le Chatelier et al., 2013). *Faecalibacterium prausnitzii* has previously shown to be depleted in 239 MetS subjects, an effect shown to be partially restored following a long-term (2 years) Mediterranean diet intervention (Haro et al., 2016). Whole grain intake and increased fibre intake as a result allows *Faecalibacterium prausnitzii* to flourish (Benus et al., 2010) and encourages the anti-inflammatory properties of *Faecalibacterium prausnitzii* (Sokol et al., 2008), which may further protect against MetS.

Limitations

There were a number of potential limitations to this study. There were a small number of male subjects in the sample, therefore these results may be applicable only to women. A *de novo* method was used to classify the OTUs into taxonomies. The *de novo* method involves creating a reference panel that is sample specific, therefore replicating our results in independent populations may be difficult to achieve. But this is true of all OTU studies. As the FFQ relies on subject reporting, the accuracy of this data have been called into question. Though in the previous chapters I have shown that reported intakes of those foods significantly associated

with hippurate, coffee, fruit and whole grains, associate with biologically plausible metabolites. If misclassification had occurred it would likely have obscured real findings and not strengthened them. Furthermore, we replicate findings from feeding studies that have shown hippurate urinary excretion to be increased following the consumption of these foods.

There were a number of advantages to this study. I had a large number of unique subjects with metabolomics profiling, dietary information and microbiome profiling. I also had access to a unique longitudinal metabolomics data in order to evaluate the influence of changes in hippurate levels on MetS risk.

Conclusion

In conclusion, hippurate, is associated with gut microbiome diversity and foods high in polyphenols. Subjects whose hippurate trajectory increased showed a reduced risk of MetS and related risk factors, largely accounted for by Shannon diversity. Microbiome OTUs/taxa related to blood hippurate and related food intakes, including increased abundances of Actinomycetaceae, *Eubacterium* and *Ruminococcus* and reduced OTUs within the order Clostridiales and the species *Faecalibacterium prausnitzii* associated with an increased MetS risk. These findings support the gut microbiome as an important mediator in the relationship between diet, in particular high polyphenol foods, and metabolic disease. The potential of hippurate as a marker of alpha-diversity and the interplay between high polyphenol foods, gut microbes and hippurate production should now be established in mechanistic studies ideally under a dietary intervention setting.

Chapter 7 Untangling the Relationship Between Diet and Visceral Fat Mass Through Blood Metabolomics and Gut Microbiome Profiling

In this chapter I created a visceral fat mass diet (VFM) score using the intakes of top foods associated with VFM. I then examined the blood metabolomics and gut microbiome profiles associated with the diet score with the aim of identifying potential biological intermediates linking diet with VFM development.

The work from this chapter has been formatted as a manuscript and submitted to the *International Journal of Obesity*.

7.1 Introduction

Increased visceral fat (VF) is a primary risk factor for cardio-metabolic diseases. Observational studies examining the impact of habitual food consumption on VFM or waist circumference (WC) have found that increased intakes of fruit (Romaguera et al., 2011), dairy products (Romaguera et al., 2011) and related nutrients (Fischer et al., 2015), and whole grains (Caron-Jobin et al., 2011) and fibre (Hairston et al., 2012, Fischer et al., 2015) are associated with lower risk, whereas higher intakes of fried foods and fat (Mollard et al., 2014, Romaguera et al., 2011), alcohol, red and processed meats (Romaguera et al., 2011) and related nutrients (Fischer et al., 2015), sugar-sweetened beverages (Ma et al., 2014, Mollard et al., 2014, Odegaard et al., 2012, Romaguera et al., 2011), refined grains (Caron-Jobin et al., 2011, Romaguera et al., 2011) and high glycaemic index foods (Dal Molin Netto et al., 2014, Romaguera et al., 2010) are associated with higher risk of VF.

Within the past decade, the effects of an unhealthy dietary pattern have been increasingly explored. For instance, Nettleton and colleagues created a diet protective health score using self-reported intakes of whole grains, fish, fruits, vegetables, nuts/seeds (favorable) and red/processed meats, sweets, sugar-sweetened beverages and fried potatoes (unfavorable). They were then able to show that the score was a powerful means to examine gene X diet interactions in obesity in 68,317 subjects of European ancestry (Nettleton et al., 2015). Similarly, Romaguera et al. (Romaguera et al., 2011) formed a summary score

combining the intake frequencies for all foods associated with changed in WC over a 5.5 year median in 48,631 men and women. However, dietary pattern scores have not yet been applied against high-throughput omic datasets.

Metabolomics is being increasingly used in dietary studies in order to further explore the mechanisms of diet on metabolic disease development as extensively shown in this thesis. A recent study from our group showed that some of the metabolites associated with VF mass relate also to type 2 diabetes, insulin resistance and blood pressure (Menni et al., 2016). Moreover, we have found that self-reported reported food intakes are associated with 106 unique metabolites (Pallister et al., 2016), which confirmed the strong role of food intake on metabolic traits (see Chapter 4). At this time, the characteristic metabolomics profile of a metabolically unhealthy diet has yet to be established, nor have those primary metabolites that connect diet to VFM development been defined.

Recent studies are alluding to a role of the gut microbiota in VF development by interacting with food compounds and as a result, supplying metabolites to our bodies (Shoaie et al., 2015), this is supported by in Chapters 4 and 5 where I showed food intakes to be correlated with microbially-derived metabolites. Moreover, I found microbiome alpha diversity and hippurate (a microbial co-metabolite) to be associated with a reduced risk of MetS supporting the interplay of diet, the microbiome and the metabolome in metabolic disease development. These interactions may also intensify the health effects of a calorific, low nutrient dense Western Diet. Through feeding rodents high-fat or high-fat/high-sugar diets, early studies have shown HF feeding increases Firmicutes and reduces Bacteriodetes (Taira et al., 2015) abundances, reduces the abundance of the class Clostridia which was additionally linked to higher visceral adipose tissue (Etxeberria et al., 2015), and increases abundances of sulfidogenic bactera which was associated with higher intestinal inflammation (Shen et al., 2014). Under high fat feeding of animals, cranberry (Anhe et al., 2015) and pomegranate (Neyrinck et al., 2013) polyphenols, resveratrol from red wine (Neyrinck et al., 2013) and gluco-oligosaccharide (Serino et al., 2012) have been found to be protective of obesity and associated inflammation in by changing the gut microbiome profiles. Also, conjugated linoleic acid, a constituent of dairy fats has been shown to encourage fat loss, which may be due to increasing the abundance of Bacteroidetes (Marques et al., 2015).

Previous studies have not undertaken a multi-omic approach to decipher the influence of an unhealthy diet on VFM, as such the aims of this chapter were:

i) To identify foods most strongly associated with VFM from these develop and validate a predictive dietary VFM-risk score.

ii) To characterise the blood metabolomics profile of the VFM-risk score.

iii) To characterise the gut microbiome profile of the dietary VFM-risk score.

7.2 Materials and methods

Figure 7-1 shows the summary of the study protocol.

For the analysis I used FFQs completed between 1995 and 2001, in 2007 and 2014 to 2015. I used the 20 food groups as defined in **Section 5.2.1**.

7.2.1 Visceral Fat Mass

During clinical visits a train research nurse or assistant determined visceral fat mass (VFM; g) in 3457 twins using Dual-Energy X-ray Absorptiometry (DXA; Hologic QDR; Hologic, Inc., Waltham, MA, USA) whole-body scanning in the supine position. To analyse the scans, the QDR System Software Version 12.6 was used. VFM was estimated at one cross-section of the whole body (L4-L5), this is the usual area of a computed tomography slice. If VFM was 4 SD above or below the mean VFM, twins were excluded from the analysis. I normalized the VFM data using a rank-based inverse-normalization as it did not follow a normal distribution.

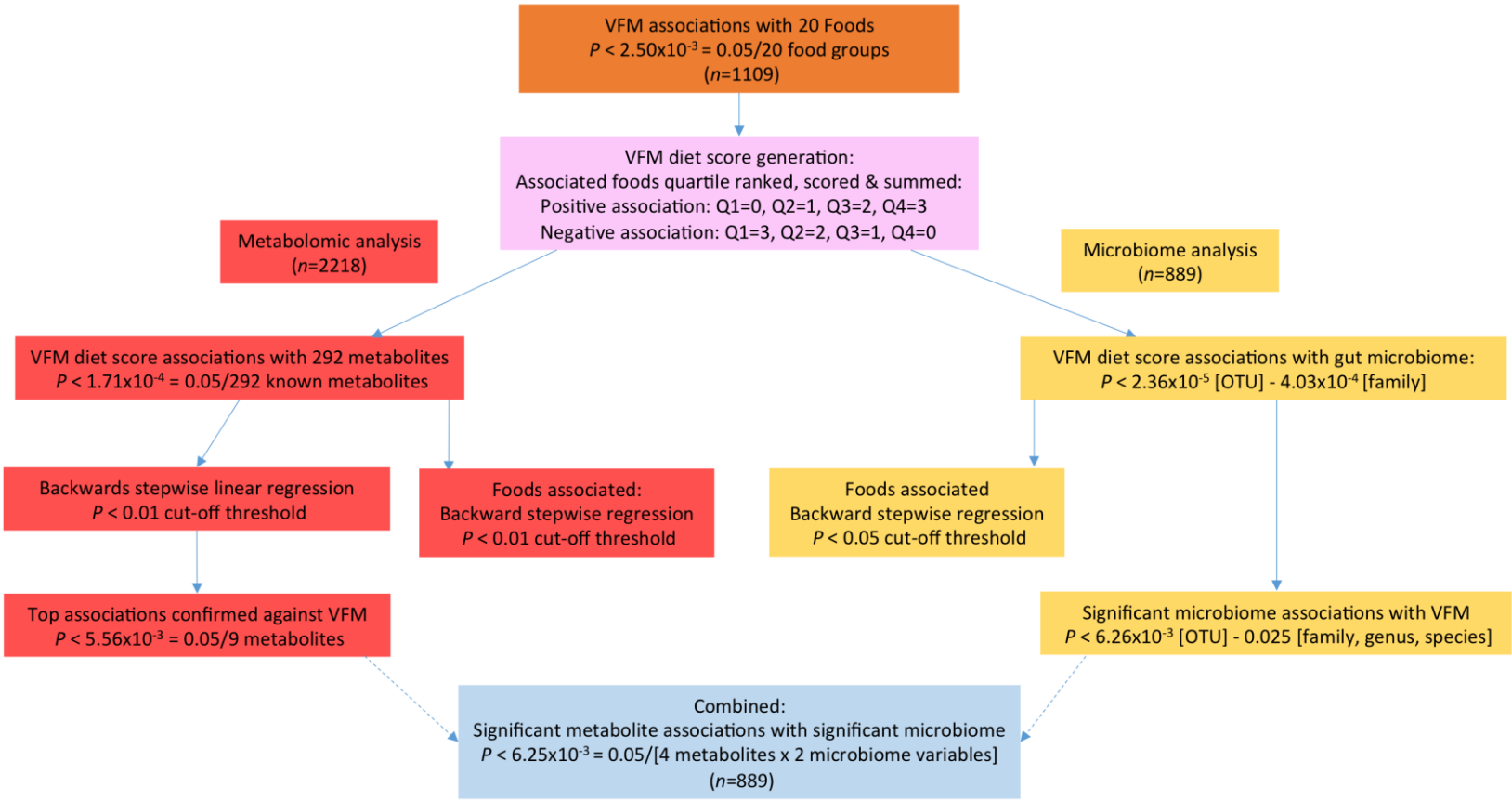
7.2.2 Metabolomic profiling

For this chapter, I used 292 chemically identified metabolites analyzed by the non-targeted Metabolon platform. Quality control of the metabolomics dataset has been described in **Section 3.1.6.1**. For the analysis I included 2218 twins (99.8% female) with metabolomics profiling, BMI and VFM data within ± 5 years of FFQ completion.

7.2.3 Gut microbiome profiles

16S rRNA gene sequencing was used to determine the faecal bacterial profiles. For a previous study (Goodrich et al., 2014b) microbial DNA was extracted, amplified, sequenced and processed by the Cornell technical team, since this earlier study an additional ~1000 samples were collected and processed using the same protocols. Quality control of the microbiome data were undertaken by Tiphaine Martin, Matthew Jackson and Dr Michelle Beaumont. At 97% sequence similarity, sequencing reads were condensed as operational taxonomic units (OTUs). To do this, UCLUST open-reference clustering was used against Greengenes v13_5 reference within QIIME 1.7.0. Of the total sequences, 6.2% did not cluster to the reference and were therefore excluded from further analysis (Goodrich et al., 2014b).

Figure 7-1. Outline of the study design



OTUs that were found in fewer than 25% of individuals were removed from the analysis. Using this threshold and after removing singletons, of 9,840 OTUs only 2,118 OTUs (16%) remained for the final analysis. All OTU counts, including those OTUs present in less than 25% of individuals, were collapsed into taxonomies at the family (124 taxa), genus (283 taxa) and species (153 taxa) levels. For the analysis I only considered taxa fully classified within each taxonomic level. Following rarefaction of the complete OTU table to 10000 reads per sample, alpha-diversity was determined by Shannon's phylogenetic diversity (Faith, 1992) using QIIME. Using linear regression, OTUs were adjusted for technical covariates which included sequencing run and sequence number per sample. I normalized the data using rank-based inverse normalization. I analyzed a subsample (*n*=889) of twins from the FFQ, VFM and metabolomics dataset that additionally had fecal microbiome profiling.

7.2.4 Muther expression data

Gene expression of abdominal fat samples in 825 individuals were analysed with the Illumina Human HT-12 V3 for the Muther study, as described previously (Grundberg et al., 2012). Using random intercept linear regression, 586 individuals were analyzed for AT expression associations with the top metabolite adjusting for age, BMI, metabolite and expression batches, and family relatedness. This analysis was conducted by Dr Cristina Menni.

7.2.5 Statistical analysis

I conducted all statistical analysis using Stata version 12.

7.2.5.1 Food group associations with VFM and diet score formation and heritability

First, I randomly allocated twins to two independent groups: the test (*n*=1109) and replication (*n*=1109) groups. I ensured that twin pairs were assigned to the same group. To identify food group associations with VFM, in the test group I ran a linear regression for each of the 20 food groups as predictors of VFM (residual adjusted for BMI), I adjusted for covariates including total fat mass, age, sex, height^2, family relatedness and DXA batch. The cut-off for statistical significance was defined using Bonferroni correction ($P<2.50\text{x}10^{-3}$ = [0.05/20 food groups]). I included food groups that were significantly associated with VFM in the risk score. I calculated the score, first I quartile ranked the consumption frequencies of associated foods and then assigned a score of 0 to 3 by the direction of the association (positive association: Q1=0, Q2=1,

Q3=2, Q4=3; negative association: Q1=3, Q2=2, Q3=1, Q4=0). Next I summed each food group score to create the final risk score that ranged from 0 to 15. I used linear structural equation modelling in Mx (Neale et al., 1992, Neale et al., 2003) to determine the heritability of the VFM diet score (consult **Section 3.2.1** for analysis details).

7.2.5.2 Binary classification test

I next tested the VFM risk score in the replication group by first fitting a logistic regression model with the diet score predicting low and high VFM. I defined the lower tertile of VFM as a negative outcome (0; *n*=369) and the upper tertile of VFM as a positive outcome (1; *n*=370). In the model I adjusted for covariates including total fat mass, age, sex, height^2, family relatedness, DXA batch, and BMI category (1: <18.5 kg/m^2; 2: ≥18.5-24.9 kg/m^2; 3: ≥25-29.9 kg/m^2; 4: ≥30 kg/m^2). To assess the predictive ability of the VFM diet score, I conducted a binary classification test. The ability of the VFM diet score to identify twins with high VFM accurately (sensitivity; true positive rate) and identify twins with low VFM accurately (specificity; true negative rate) was predicted. The receiver operating characteristic curve (ROC) was then generated by plotting the true positive rate against the false positive rate at a number of threshold settings.

7.2.5.3 VFM diet score and metabolomics and microbiome

For each metabolite or microbiome taxon, I ran a random intercept linear regression analysis for each metabolite or microbiome OTU/taxa adjusting for age, BMI, sex and metabolite batch (for metabolomics) or Shannon Index (for microbiome), and family relatedness:

$$Y_i = \beta_0 + \boldsymbol{\beta_i X_{ij}} + \delta_{1i} age_{ij} + \delta_{2i} BMI_{ij} + \delta_{3i} Sex_{ij} + \gamma_{1i} Z_{ij} + \zeta_j + \varepsilon_{ij}$$

where Y_i is the metabolite/microbiome OTU/taxa, X_{ij} is the VFM diet score of twin *j* from pair *i*, Z_{ij} is the metabolite batch or Shannon Index and ζ_j, is the family-specific error component that captures the unobserved heterogeneity or family characteristics. I also used Bonferroni correction to account for multiple testing giving a significance threshold for metabolites of 1.71×10^{-4} (0.05/(292 known metabolites)). For the microbiome analysis I defined statistical significance within each taxonomic level, **Table 7-1** shows the thresholds by level.

A large number of potentially correlated metabolites were associated with the VFM diet, therefore I next undertook a backwards stepwise linear regression including all associated metabolites, using a cut-off threshold of *P*<0.01.

Table 7-1. Statistical significance thresholds for microbiome analysis using Bonferroni correction

	VFM diet score		VFM	
Level [(1)]	**Number of variables**	***P***	**Number of variables [(2)]**	***P***
Family	124	4.03×10^{-4}	2	0.025
Genus	283	1.77×10^{-4}	2	0.025
Species	153	3.27×10^{-4}	2	0.025
OTU	2118	2.36×10^{-5}	8	6.25×10^{-3}

(1) Bonferroni correction was calculated within each level.
(2) Variables significantly associated with the VFM diet score were tested for their association against VFM.

7.2.5.4 VFM and metabolomics and microbiome

To determine if the remaining metabolites, and OTUs/taxa (both residual-adjusted for BMI) associated with the VFM diet score were also associated with VFM, I used each of these variables as predictors of VFM (residual-adjusted for BMI at scan) in a linear regression including the covariates scan age, sex, total fat mass, height^2, scan batch and metabolite batch (for metabolomics) or Shannon Index (for microbiome), family relatedness and for the VFM diet score.

$$VFM_i = \beta_0 + \boldsymbol{\beta_{1i} Yi} + \beta_{2i} X_{ij} + \delta_{1i} age_{ij} + \delta_{3i} Sex_{ij} + \delta_{4i} Total\_fat\_mass_{ij} + \delta_{5i} Height\_squared_{ij} + \delta_{6i} Scan\_batch_{ij} + \gamma_{1i} Z_{ij} + \zeta_j + \varepsilon_{ij}$$

where Y_i is the metabolite/taxon, X_{ij} is the VFM diet score of twin j from pair i, Z_{ij} is the metabolite batch for metabolomics data or Shannon Index for microbiome data and ζ_j, is the family-specific error component that captures the unobserved heterogeneity or family characteristics. Associations which passed the Bonferroni cut-off were considered significant. For metabolites the threshold was 5.56×10^{-3} (0.05/9 metabolites). The assignment for statistical significance for the microbiome analysis is shown in **Table 7-1**.

I identified MZ twins who were discordant (1 SD apart) for VFM, this included 80 pairs in the metabolomics dataset and 27 pairs in the microbiome subsample. In this group, I ran a linear regression to validate top associations between diet, metabolites and the microbiome and VFM.

7.2.5.5 Top microbiome and metabolite associations with food groups

To confirm the associations between the metabolites/microbiome and the VFM diet score were not due to intakes of other foods, I ran a backward stepwise linear regression model including the VFM diet score and food groups not forming the score to predict the associated metabolite and OTU/taxa. I used a cut-off threshold of $P<0.01$ for metabolites and $P<0.05$ for microbiome.

I then assessed whether the foods contributing to the VFM diet score were independently driving the association between the VFM diet score and taxa/metabolites. I fitted all 20 food groups into a backward stepwise linear regression model using cut-off thresholds of $P<0.01$ for metabolites and $P<0.05$ for microbiome.

7.2.5.6 VFM diet score association with VFM mediated by metabolite and microbiome

I first determined the proportion of the variance of VFM attributable to the VFM diet score after accounting for all covariates (age, sex, BMI, total fat, height^2, family relatedness, metabolite batch, Shannon Index and scan batch). This quantity was represented by r^2_x. I next determined the proportion of the variance in VFM explained by the VFM diet score by adjusted for the same covariates as above but additionally including the metabolite or OTU/taxon (r^2_{xy}). To calculate the percentage of the VFM diet score association that was mediated by the metabolite or OTU/taxa (r^2_y), I determined the proportion of the variance of VFM that is due to the VFM diet score association with the metabolite or OTU/taxa: $1 – (r^2_{xy}/r^2_x)$.

7.2.5.7 Microbiome association with VFM mediated by metabolite

I determined the proportion of the variance of VFM attributable to the taxon after taking into account all covariates as above in addition to the VFM diet score (r^2_x). I next determined the proportion of the variance in VFM explained by the taxon after adjusting for the same covariates as above but additionally adjusting for the metabolite (r^2_{xy}). To calculate the percentage of the taxon association that was mediated by the metabolite (r^2_y), I determined the proportion of the variance of VFM that is due to the metabolite association with the taxon: $1 – (r^2_{xy}/r^2_x)$.

7.3 Results

7.3.1 VFM food group associations

The characteristics of the sample as well as the training and validation groups are shown in **Table 7-2**.

Table 7-2. Study population characteristics

	Whole (*n*=2218)	Training (*n*=1109)	Validation (*n*=1109)
	Mean (SD)	Mean (SD)	Mean (SD)
Age (years)	58.3 (10.9)	58.1 (11.1)	58.4 (10.8)
BMI (kg/m^2)	26.2 (4.8)	26.1 (4.8)	26.4 (4.8)
Total fat (grams)	6827 (2180)	6779 (2178)	6911 (2258)
Visceral fat (grams)	562.7 (293.7)	553.9 (293.8)	572.7 (295.5)
Height (m)	1.61 (0.06)	1.62 (0.06)	1.62 (0.06)
Sex (M:F)	4:2214	0:1109	4:1105
Food intakes (servings/week)			
Vegetables	34.3 (15.4)	33.8 (15.0)	34.8 (15.8)
Fruit and fruit juices	22.4 (13.3)	22.7 (13.8)	22.1 (12.8)
Nuts and legumes	7.9 (5.5)	7.8 (5.6)	8.0 (5.4)
Whole grains	10.2 (7.9)	10.2 (7.7)	10.2 (8.1)
Refined grains	8.4 (7.4)	8.4 (7.4)	8.5 (7.4)
Fermented dairy	6.3 (4.8)	6.4 (4.7)	6.3 (4.8)
White meat	1.9 (1.3)	1.9 (1.3)	2.0 (1.3)
Seafood	2.4 (2.1)	2.4 (1.9)	2.5 (2.2)
Red, processed meat and eggs	6.9 (4.0)	6.9 (3.9)	6.9 (4.0)
Fried and fast foods	4.5 (3.4)	4.5 (3.4)	4.6 (3.4)
Sweets and sweet baked products	15.7 (13.6)	16.2 (13.3)	15.2 (13.8)
Chocolate	4.0 (5.8)	4.1 (6.0)	3.9 (5.7)
Butter and cream	4.4 (6.6)	4.7 (6.8)	4.1 (6.4)
Spreads and dressings	8.3 (8.4)	8.4 (8.8)	8.2 (8.1)
Soy and other milks	0.2 (0.8)	0.2 (0.8)	0.2 (0.8)
Milk	3.9 (4.2)	3.8 (4.1)	4.0 (4.3)
Soda	2.0 (5.2)	1.9 (5.5)	2.1 (4.9)
Coffee	8.7 (10.4)	8.8 (10.5)	8.6 (10.4)
Tea	19.3 (13.7)	19.5 (13.8)	19.1 (13.6)
Alcohol	6.0 (8.0)	5.7 (7.8)	6.3 (8.3)

No significant differences between the training and validation sets.

I found 5 food groups were significantly associated with VFM in the training dataset (**Table 7-3**). Higher intakes of fruit, whole grain and fermented dairy products were associated with reduced VFM, whereas increased intakes of fried and fast foods and red, processed meat and eggs were associated with increased VFM.

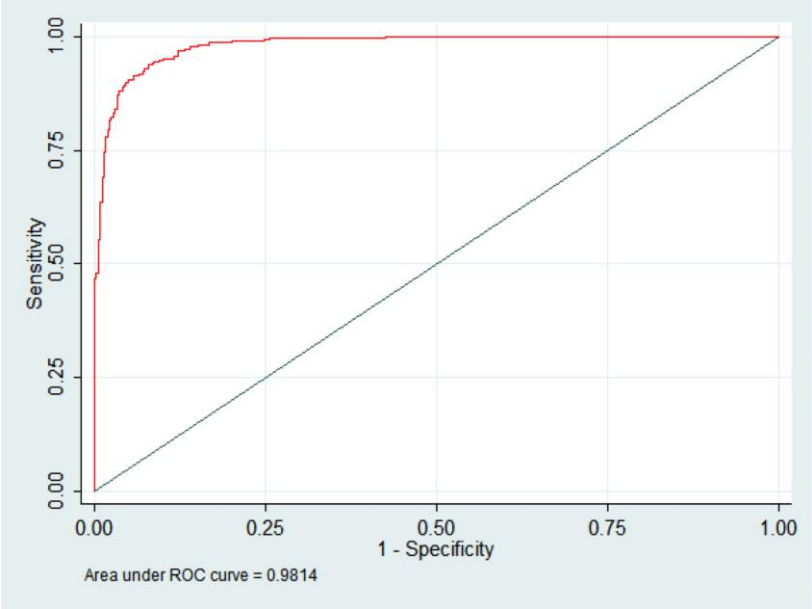
Table 7-3. List of food groups significantly associated with VFM and the VFM diet score association with VFM in the training group [(1)]

Food group	Beta	SE	*P*
Fruits	-0.005	0.001	$1.95x10^{-5}$
Red, processed meat and eggs	0.016	0.005	$3.94x10^{-4}$
Fermented dairy products	-0.011	0.004	$1.14x10^{-3}$
Fried and fast foods	0.015	0.005	$1.18x10^{-3}$
Whole grains	-0.008	0.002	$1.27x10^{-3}$
VFM diet score [(2)]	0.033	0.006	$7.38x10^{-9}$

(1) 20 food groups were tested for their association with VFM in the training group adjusting for covariates (scan batch, age, BMI, total fat, height^2, sex and family relatedness) and multiple testing (P<0.0025).
(2) The VFM diet score was formed from quartile ranking the VFM-associated food groups, scoring them according to the direction of association and summing the scores.

Based on the direction of the food group association with VFM in the training dataset, I created the VFM diet score and tested it in the validation sample. I found 93.21% of the subjects were accurately classified into low and high VFM by the diet score. Moreover, the sensitivity (true positive rate) and specificity (true negative rate) of the VFM diet score were 93.72% and 92.70%, respectively. The ROC curve is displayed in **Figure 7-2.** The area under the receiver operating curve (AUC) was 0.9841 (95% CI: 0.9772; 0.9911). Notably, I found the association between the diet score and VFM to be significant in the VFM-discordant MZ twin sample (0.281[0.091]; P=0.002).

Figure 7-2. Receiver operating characteristic curve for the VFM diet score ability to predict the bottom and top tertiles of VFM



The best-fitting model for the heritability analysis of the VFM diet score was the AE model, with a heritability estimate of 44% (95% CI: 37%, 50%) **(Table 7-4** shows details of heritability analysis).

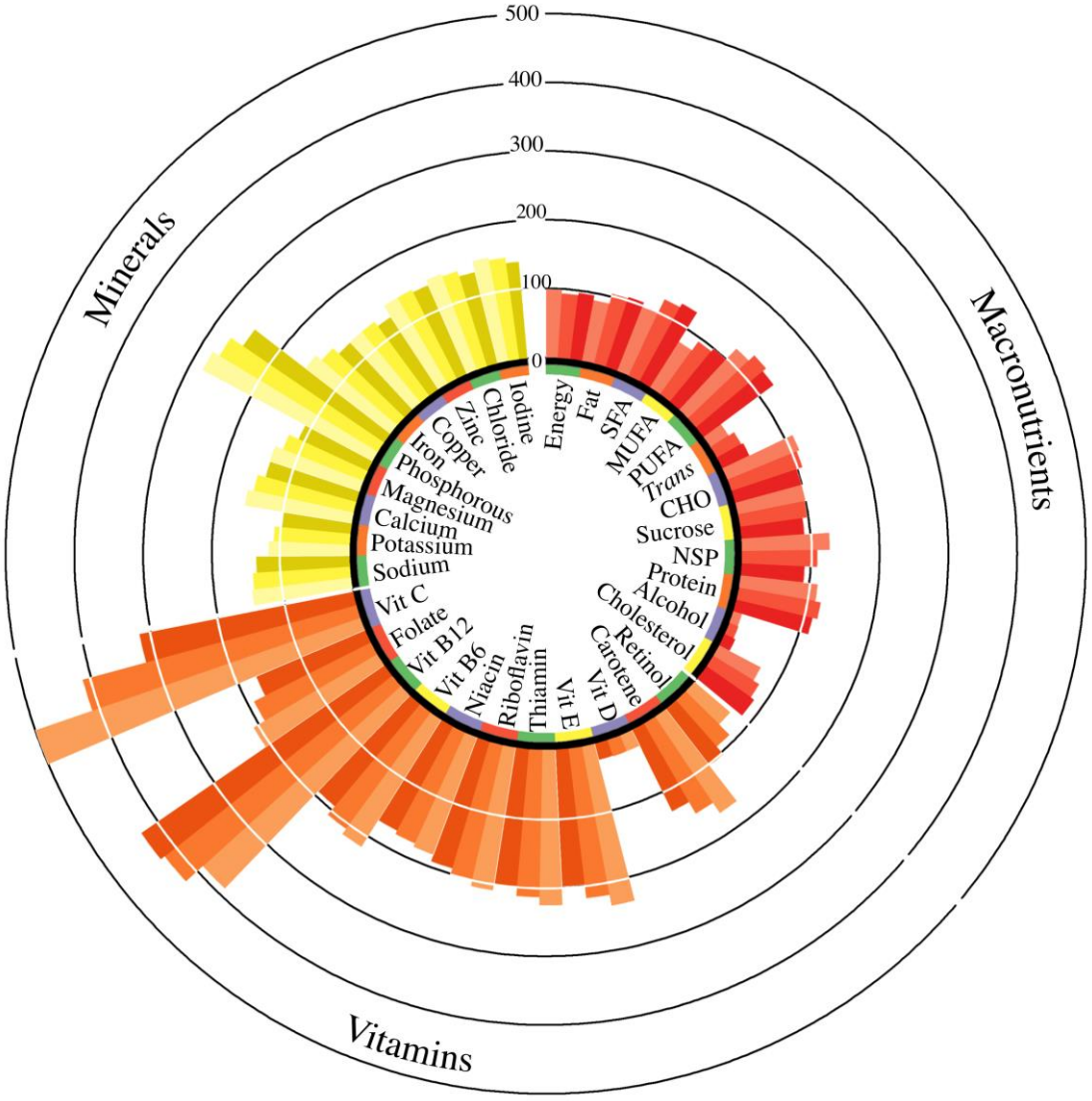
Table 7-4. Results of the structural equation modelling for heritability of the visceral fat mass diet score

Model	A	C	E	log-likelihood	X^2	*P*	AIC
ACE	0.44 (0.31, 0.50)	0.00 (0.00, 0.11)	0.56 (0.50, 0.63)	9067			5493
AE*	0.44 (0.37, 0.50)	-	0.56 (0.50, 0.63)	9067	0.00	Incalc	5491
CE	-	0.33 (0.27, 0.38)	0.67 (0.62, 0.73)	9090	23.40	0.000	5514
E	-	-	1.00 (1.00, 1.00)	9192	124.70	0.000	5613

*Best fitting model; Abbreviations: A, additive; C, common; E, environmental; df, degrees of freedom; AIC, Akaike's information criterion; Incalc, incalculable.
Structural equation modeling was performed on the visceral fat mass diet score. The best-fitting model is indicated by the lowest AIC.

Figure 7-3 displays the nutrient profile of the VFM diet score (**consult Appendix F Table 1 for details**). Individuals scoring highly on the VFM diet score had significantly increased intakes of fat (notably monounsaturated fatty acids) and reduced carbohydrate (notably, sucrose and non-starch polysaccharides [NSP]), vitamin C and magnesium intakes.

Figure 7-3. Nutrient profile of the VFM diet score presented as percentages of the UK dietary reference values by tertile of the VFM diet score



Average nutrient intakes by increasing tertile of the VFM diet score from clockwise (lightest to darkest) were assessed for percentage of the recommended intakes for 55-year-old women (according to the UK Dietary Reference Values (Health., 1991)). Using VFM diet score by tertile as the predictor of the residual energy adjusted nutrient intakes in a linear regression statistically significant trends (P<0.001) were observed for all nutrients, except polyunsaturated fatty acids, protein, zinc and vitamin D. Carotene and retinol are represented as percentage of the recommended intake for total retinol equivalents. There is no UK DRV for vitamin D therefore 10 ug/d was used. Abbreviations: SFAs, saturated fatty acids; MUFAs, monounsaturated fatty acids; PUFAs, polyunsaturated fatty acids; *Trans*, *trans* fatty acids; CHO, carbohydrates; NSP, non-starch polysaccharides; vit, vitamin.

7.3.2 VFM diet score metabolomics associations

Thirty metabolites were significantly associated (P<1.71x10^{-4}) with the VFM diet, shown in **Table 7-5**. Overall, these metabolites belonged to 21 unique sub-pathways, specifically, there were associations with 12 lipids, 5 amino acids, 5 vitamins and cofactors, 3 carbohydrates, 3 xenobiotics, 1 nucleotide and 1 peptide.

After I adjusted for other food group intakes (**Table 7-5**) all associations between the VFM diet score and metabolites remained (P<0.01) however 6 associations did not pass adjustment for multiple testing (primarily lysolipids).

7.3.3 Metabolites associated with the VFM diet score and food groups independently

Following the backward stepwise regression including all food groups, I found that 18 metabolites were significantly (P<3.33×10^{-4} (0.05/[5 food groups x 30 metabolites])) associated with the food groups forming the VFM diet score (**Table 7-5**). Fruit intakes were associated with 11 metabolites, suggesting it was an important influence on associations between the VFM diet score and metabolites. Moreover, 5 metabolites were significantly associated with whole grain intake, 3 metabolites with red, processed meat, and eggs, and 2 metabolites with fried and fast food intakes. I found associations between the food groups forming the VFM diet score and the 12 remaining metabolites did not pass statistical significance.

Table 7-5. List of metabolites significantly associated with the VFM diet score and with each food group independently (*P*<0.01 backward stepwise regression)

		VFM Score [1]		VFM Score adjusted foods [2]		Foods associated [3]	
Metabolite	**Sub-pathway**	**Beta (SE)**	***P***	**Beta (SE)**	***P***	***P*<0.01**	***P*<3.33x10^{-4}**
Eicosapentaenoate	Essential fatty acid	-0.058(0.007)	3.24x10^{-17}	-0.052(0.006)	2.22x10^{-15}	Fruit (0.005(0.002)) FD (0.011(0.004)) FF (-0.017(0.006))	WG (0.011(0.003))
Indolepropionate	Tryptophan metabolism	-0.056(0.007)	1.58x10^{-16}	-0.054(0.007)	2.91x10^{-16}	Fruit (0.006(0.002)) WG (0.008(0.003)) RM (-0.018(0.005))	
3-Carboxy-4-methyl-5-propyl-2-furanpropanoate	Fatty acid, dicarboxylate	-0.055(0.007)	2.43x10^{-16}	-0.045(0.006)	3.19x10^{-12}	FF (-0.023(0.007))	WG (0.013(0.003))
Docosahexaenoate	Essential fatty acid	-0.054(0.007)	6.17x10^{-16}	-0.045(0.006)	1.30x10^{-12}	WG (0.010(0.003))	Fruit (0.008(0.002))
Stachydrine	Food component, Plant	-0.056(0.007)	7.95x10^{-16}	-0.057(0.007)	2.59x10^{-16}	RM (-0.019(0.006))	Fruit (0.014(0.002))
3-Phenylpropionate	Phenylalanine & tyrosine metabolism	-0.059(0.007)	1.70x10^{-15}	-0.053(0.008)	3.53x10^{-12}	Fruit (0.006(0.002))	WG (0.014(0.003)) FF (-0.028(0.007))
Hippurate	Benzoate metabolism	-0.052(0.007)	1.84x10^{-13}	-0.052(0.007)	5.06x10^{-14}		Fruit (0.010(0.002)) WG (0.011(0.002))
Catechol sulfate	Benzoate metabolism	-0.051(0.007)	1.75x10^{-12}	-0.046(0.007)	1.94x10^{-10}	WG (0.008(0.003))	Fruit (0.008(0.002))
Glycerate	Glycolysis, gluconeogenesis, pyruvate metabolism	-0.049(0.007)	3.81x10^{-12}	-0.042(0.007)	1.57x10^{-9}		Fruit (0.010(0.002))
Pyridoxate	Vitamin B6 metabolism	-0.049(0.007)	9.08x10^{-12}	-0.042(0.007)	2.96x10^{-9}		Fruit (0.007(0.002)) WG (0.011(0.002))
Threitol	Nucleotide sugars, pentose metabolism	-0.044(0.007)	7.73x10^{-11}	-0.043(0.007)	1.65x10^{-10}		Fruit (0.011(0.002))
Butyrylcarnitine	Fatty acid metabolism (also BCAA metabolism)	0.040(0.007)	4.77x10^{-9}	0.039(0.007)	9.09x10^{-9}		RM (0.020(0.005))
alpha-Hydroxyisovalerate	Valine, leucine and isoleucine metabolism	0.037(0.007)	1.12x10^{-7}	0.027(0.006)	4.48x10^{-5}		RM (0.022(0.005))
1-Arachidonoylglycerophospho-ethanolamine*	Lysolipid	0.038(0.007)	2.09x10^{-7}	0.030(0.007)	3.43x10^{-5}	WG (-0.011(0.003)) FD (-0.013(0.005))	
Threonate	Ascorbate and aldarate metabolism	-0.036(0.007)	4.51x10^{-7}	-0.033(0.007)	3.66x10^{-6}		Fruit (0.007(0.002))
1,5-Anhydroglucitol (1,5-AG)	Glycolysis, gluconeogenesis, pyruvate metabolism	0.035(0.007)	1.80x10^{-6}	0.026(0.007)	3.46x10^{-4}		Fruit (-0.008(0.002))
Uridine	Pyrimidine metabolism, uracil containing	-0.032(0.007)	2.05x10^{-6}	-0.03(0.007)	9.21x10^{-6}	WG (0.008(0.003))	
Stearidonate	Long chain fatty acid	-0.033(0.007)	2.65x10^{-6}	-0.028(0.007)	3.20x10^{-5}	None	
1-Docosahexaenoylglycerophosphocholine*	Lysolipid	-0.032(0.007)	2.95x10^{-6}	-0.027(0.007)	4.14x10^{-5}	FD (0.014(0.004))	

Table 7-5. List of metabolites significantly associated with the VFM diet score and with each food group independently (*P*<0.01 backward stepwise regression)

		VFM Score [1]		VFM Score adjusted foods [2]		Foods associated [3]	
Metabolite	**Sub-pathway**	**Beta (SE)**	***P***	**Beta (SE)**	***P***	***P*<0.01**	***P*<3.33x10^{-4}**
Bilirubin (Z,Z)	Hemoglobin and porphyrin metabolism	-0.033(0.007)	3.88x10^{-6}	-0.032(0.007)	4.69x10^{-6}		FF (-0.020(0.006))
1-Oleoylglycerophosphoethanolamine	Lysolipid	0.032(0.007)	9.21x10^{-6}	0.023(0.007)	1.35x10^{-3}	WG (-0.008(0.003))	
1-Arachidonoylglycerophosphocholine*	Lysolipid	0.032(0.007)	1.76x10^{-5}	0.024(0.007)	1.26x10^{-3}	WG (-0.010(0.003))	
Pantothenate	Pantothenate and CoA metabolism	-0.029(0.007)	4.20x10^{-5}	-0.029(0.007)	6.28x10^{-5}	None	
4-Androsten-3beta,17beta-diol disulfate 1*	Sterol, Steroid	0.032(0.008)	4.67x10^{-5}	0.019(0.007)	7.86x10^{-3}	None	
X-11793--oxidized bilirubin*	Hemoglobin and porphyrin metabolism	0.028(0.007)	5.10x10^{-5}	0.027(0.007)	7.01x10^{-5}	WG (-0.009(0.002)) FD (-0.015(0.005))	
trans-4-Hydroxyproline	Urea cycle; arginine-, proline-, metabolism	0.027(0.007)	8.28x10^{-5}	0.031(0.007)	7.13x10^{-6}	FD (-0.013(0.005))	RM (0.028(0.006))
Glycoursodeoxycholate	Bile acid metabolism	0.032(0.008)	1.00x10^{-4}	0.028(0.008)	5.70x10^{-4}	None	
gamma-Glutamylvaline	gamma-glutamyl	0.023(0.006)	1.11x10^{-4}	0.023(0.006)	1.38x10^{-4}	FD (-0.014(0.004))	Fruit (-0.005(0.001))
1-Eicosatrienoylglycerophosphocholine*	Lysolipid	0.028(0.007)	1.32x10^{-4}	0.020(0.007)	4.40x10^{-3}	WG (-0.009(0.003))	
Proline	Urea cycle; arginine-, proline-, metabolism	0.028(0.007)	1.44x10^{-4}	0.028(0.007)	1.45x10^{-4}	FF (0.015(0.006))	Fruit (-0.008(0.002))

FD: Fermented dairy; FF: Fried and fast foods; RM: Red meat; WG: Whole grain products

(1) Metabolite associations with the VFM diet score were adjusted for covariates (batch effects, age, BMI and sex) and multiple testing (*P*<1.71x10^{-4}).

(2) The VFM diet score and 15 food groups not forming the score were fitted into a backward stepwise linear regression model to predict each significant metabolite using *P*<0.01 as the cut off threshold.

(3) All 20 food groups were fitted into a backward stepwise linear regression model to predict each significant metabolite using *P*<0.01 as the cut off threshold. Significant results shown only for foods forming the VFM diet score. Associations passing the Bonferonni cut-off were considered statistically significant: *P*<3.33x10^{-4} = 0.05/[5 food groups x 30 metabolites].

7.3.4 Metabolites associated to both the VFM diet and VFM

I included all 30 metabolites in a backward stepwise linear regression and found 9 metabolites were significantly and independently associated with the VFM diet score (**Table 7-6**), these metabolites accounted for 14% of the variance in the VFM diet score. Notably, four of these metabolites also passed the cut-off four multiple testing ($P<5.56x10^{-3}$). I also found 6 out of the 9 metabolites were nominally associated ($P<0.05$) with both the VFM diet and VFM. The remaining 3 metabolites were associated only with the VFM diet and not VFM.

I found higher scores on the VFM diet score and high VFM (independently of diet and total body fat) were associated with lower hippurate and bilirubin (Z,Z), and higher alpha-hydroxyisovalerate and butyrylcarnitine (**Table 7-6**). Notably, in the VFM-discordant MZ twins, associations between VFM and butyrylcarnitine (0.199[0.087]; $P=0.023$) and hippurate (-0.297[0.095]; $P=0.002$) were significant (**Figure 7-4**; **Appendix G Table 2**). On average, these four metabolites accounted for 18.5% (range: 13.5% [alpha-hydroxyisovalerate]-28.9% [hippurate]) of the variance in the association between the VFM diet score and VFM (**Table 7-6**).

Table 7-6. List of metabolites independently associated with the VFM diet score (*P*<0.01 in backward linear regression), their association with VFM and the proportion of the association of the VFM diet score with VFM that is mediated by the VFM diet score association with the metabolites (*P*<5.56x10^{-3})

	VFM diet score stepwise (1)		VFM (2)				
Metabolite name	beta(SE)	*P*	beta(SE)	*P*	diet R^2 no metabolite (3)	diet R^2 with metabolite (4)	% association through metabolite
Hippurate	-0.45(0.10)	2.15x10^{-5}	-0.081(0.012)	1.33x10^{-11}	0.0312	0.0222	28.8%
alpha-Hydroxyisovalerate	0.38(0.10)	9.60x10^{-5}	0.050(0.013)	1.65x10^{-4}		0.0270	13.5%
Butyrylcarnitine	0.33(0.10)	8.54x10^{-4}	0.072(0.013)	5.86x10^{-8}		0.0267	14.4%
Bilirubin (Z,Z)	-0.31(0.10)	1.76x10^{-3}	-0.049(0.013)	1.88x10^{-4}		0.0258	17.3%
Indolepropionate	-0.33(0.11)	2.21x10^{-3}	-0.030(0.012)	1.40x10^{-2}			
1-Arachidonoylglycerophospho-choline*	0.27(0.10)	5.20x10^{-3}	0.031(0.012)	1.07x10^{-2}			
Eicosapentaenoate (EPA; 20:5n3)	-0.75(0.10)	1.13x10^{-13}	0.020(0.012)	*NS*			
Threonate	-0.32(0.11)	2.59x10^{-3}	-0.016(0.012)	*NS*			
X-11793--Oxidized bilirubin*	0.33(0.11)	2.63x10^{-3}	-0.004(0.012)	*NS*			

NS= not significant: *P*>0.05

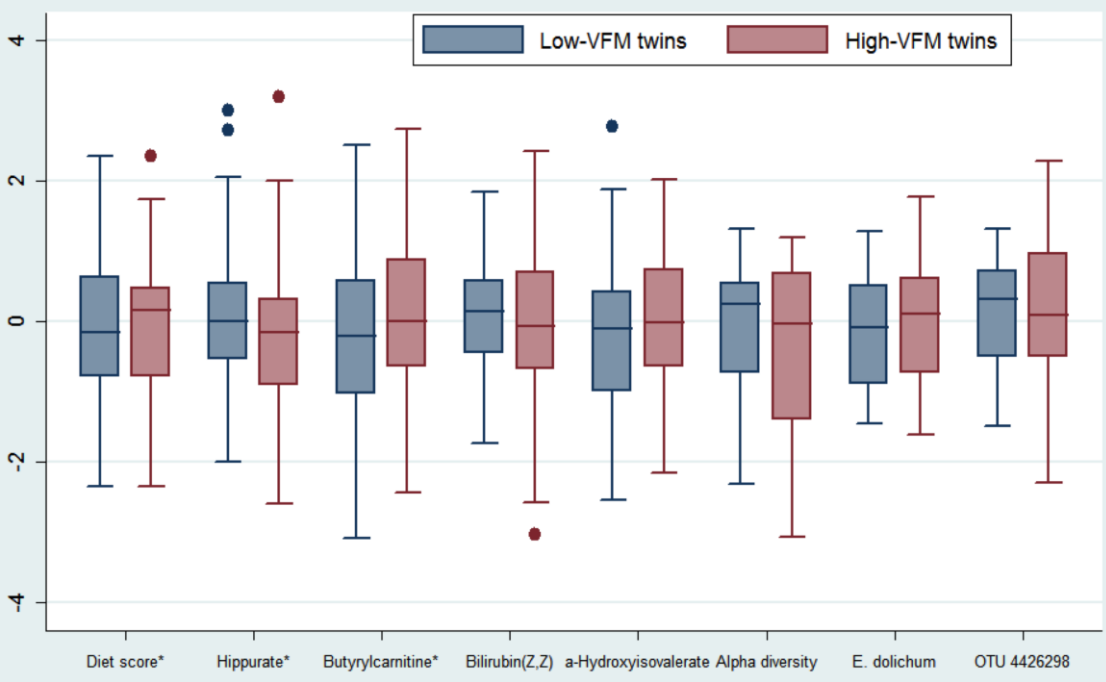
(1) Thirty metabolites significantly associated with the VFM diet score (Table 7-5) were adjusted for covariates (batch effects, age, BMI and sex) and fitted into a backward stepwise linear regression to predict the VFM diet score using *P*<0.01 as the threshold cut-off.

(2) Nine metabolites independently associated with the VFM diet score were tested for their association with VFM adjusted for covariates (age, batch effects, BMI, total fat, sex, height^2, and family relatedness). Associations passing the Bonferonni cut-off were considered significant (*P*<5.56x10^{-3}).

(3) the proportion of the variance in VFM explained by the VFM diet score after taking into account all covariates (age, sex, BMI, height^2, and batch effects).

(4) the proportion of the variance in VFM explained by the VFM diet score after taking into account all covariates as in (1) and adjusting for the metabolite.

Figure 7-4. Comparisons of the VFM diet score, alpha diversity and top microbiome and metabolite associations in the low and high MZ VFM-discordant twins



All variables were standardized to have mean=0, SD=1. A linear regression was conducted using the VFM diet score, alpha diversity (Shannon Index) and top microbiome and metabolite associations to predict VFM in the MZ discordant (1 SD apart in VFM) twin sample. Significantly (*P*<0.05) higher VFM diet scores and butyrylcarnitine, and lower hippurate were observed with increasing VFM (*).

7.3.5 VFM diet score microbiome associations

In a subsample of 889 twins I found higher scores on the VFM diet were associated with lower gut microbiome diversity (Shannon Index; -0.025[0.009], P=6.26x10^{-3}), after I adjusted for VFM this association was reduced but remained significant (-0.020[0.010], P=0.035).

I found the VFM diet score was associated significantly with 8 OTUs (**Table 7-7**) and 6 taxa (**Table 7-8**). However, after I adjusted for intakes of other food groups, the associations became nominally significant (*P*<0.05) (**Table 7-8**).

7.3.6 Microbiome taxa associated to both the VFM diet and VFM

I found higher abundances of the species species *Eubacterium dolichum* (0.057[0.019], P=2.73x10^{-3}) was associated significantly with increased VFM and a *Bifidobacterium* OTU (OTU ID: 4426298; -0.046[0.016], P=0.005) with reduced VFM (adjusting for the VFM diet score). I found that 16.4% of the effect of the VFM diet score on VFM (r^2_x= 0.0238) was mediated by *E. dolichum* (r^2_{xy}= 0.0199) and 17.2% by the *Bifidobacterium* OTU.

7.3.7 Eubacterium dolichum and hippurate associated with both VFM and VFM diet

I tested associations between those 4 metabolites significantly associated to both VFM and the VFM diet to determine if *E. dolichum* and the *Bifidobacterium* OTU may be related to the metabolomics profile. I found that higher abundances of *E. dolichum* were associated with lower levels of hippurate at the nominal level (P<0.05) after adjusting for VFM, the VFM diet score, Shannon Index and covariates (-0.075[0.032], P=0.021). Moreover, I found that 36.9% of the effect of *Eubacterium dolichum* on VFM (r^2_x= 0.0065) was mediated by hippurate (r^2_{xy}= 0.0041) independently of diet and covariates.

7.3.8 Hippurate association with adipose tissue transcriptome

Adipose tissue gene expression levels in the greater twin population, in order to provide potential mechanisms of the effect of hippurate on fat mass development. Higher levels of hippurate were associated with increased expression of neuroglobin (0.016[0.004], P=9.82x10^{-6}). Neuroglobin has roles in cellular energy maintenance as a vertebrate globin family member.

Table 7-7. List of OTUs associated with the VFM diet score (unadjusted and adjusted for other food intakes), their association with foods forming the VFM diet score and their independent association with the VFM diet score (*P*<0.05 in backward linear regression)

OTU ID	Assigned Taxonomy	VFM score [1]		VFM score adjusted foods [2]		Foods associated [3]
		Beta(SE)	*P*	Beta(SE)	*P*	*P*<0.05
4426298 [4]	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Bifidobacteriales; f__Bifidobacteriaceae; g__Bifidobacterium; s__	-0.058(0.011)	$6.19x10^{-7}$	-0.053(0.011)	$9.56x10^{-7}$	FD (0.029(0.008))* FF (-0.022(0.010))
183686	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__; s__	0.053(0.011)	$1.25x10^{-6}$	0.042(0.011)	$9.01x10^{-5}$	RM (0.018(0.009)) WG (-0.016(0.004))*
592616	k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__; s__	0.050(0.011)	$3.35x10^{-6}$	0.042(0.010)	$4.48x10^{-5}$	RM (0.027(0.009))*
2368865	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__	-0.047(0.011)	$1.11x10^{-5}$	-0.044(0.010)	$2.40x10^{-5}$	FD (0.031(0.006))*
509709	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__	-0.051(0.011)	$1.15x10^{-5}$	-0.047(0.011)	$2.57x10^{-5}$	RM (0.011(0.004)) WG (-0.023(0.008))
New.0.Reference OTU51	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Rikenellaceae; g__; s__	0.049(0.011)	$1.44x10^{-5}$	0.045(0.011)	$3.87x10^{-5}$	WG (-0.011(0.004))
3801267	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Veillonella; s__parvula	-0.043(0.011)	$1.75x10^{-5}$	-0.039(0.010)	$7.29x10^{-5}$	Fruit (0.006(0.002))
2407149	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__	-0.043(0.010)	$2.24x10^{-5}$	-0.028(0.010)	0.006	None

*= statistically significant: *P*<0.0025; FD: fermented dairy; FF: fried and fast foods; RM: red meat; WG: whole grain products

(1) OTU associations with the VFM diet score were adjusted for covariates (age, Shannon Index, BMI and sex) and multiple testing (*P*<$2.36x10^{-5}$ [Bonferroni: 0.05/2118 OTUs]).

(2) The VFM diet score and 15 food groups not forming the score were fitted into a backward stepwise linear regression model to predict each significant OTU using *P*<0.05 as the cut off threshold.

(3) All 20 food groups were fitted into a backward stepwise linear regression model to predict each significant OTU using *P*<0.05 as the cut off threshold. Significant results shown only for foods forming the VFM diet score.

(4) OTU 4426298 is the only taxon associated with VFM independently of the VFM diet score (Beta[SE]: (-0.046[0.016]; *P*=0.005).

Table 7-8. List of taxa associated with the VFM diet score (unadjusted and adjusted for other food intakes), their association with foods forming the VFM diet score and their independent association with the VFM diet score (*P*<0.05 in backward linear regression)

		VFM Score [1]		VFM Score adjusted foods [2]		Foods associated [3]
Taxon	**Level**	**beta(SE)**	***P***	**beta(SE)**	***P***	***P*<0.05**
Actinomyces	genus	0.052(0.011)	$9.77x10^{-7}$	0.052(0.011)	$9.77x10^{-7}$	FF (0.028(0.009))* RM (0.027(0.008))*
Lachnospira	genus	-0.045(0.009)	$2.79x10^{-6}$	-0.038(0.010)	$8.33x10^{-5}$	Fruit (0.006(0.002))
Actinomycetaceae	family	0.043(0.011)	$5.47x10^{-5}$	0.043(0.011)	$5.47x10^{-5}$	FF (0.021(0.010)) RM (0.024(0.008))
Eubacterium dolichum [4]	**species**	**0.042(0.011)**	**$8.47x10^{-5}$**	**0.043(0.011)**	**$6.19x10^{-5}$**	**WG (-0.010(0.004))**
Veillonella dispar	species	-0.039(0.011)	$3.05x10^{-4}$	-0.031(0.011)	$4.00x10^{-3}$	None
Anaeroplasmataceae	family	-0.037(0.010)	$3.75x10^{-4}$	-0.036(0.010)	$3.37x10^{-4}$	Fruit (0.007(0.003)) WG (0.011(0.004))

*= statistically significant: *P*<0.0025: FF: Fried and fast foods; RM: Red meat; WG: Whole grain products

(1) Taxa associations with the VFM diet score were adjusted for covariates (age, Shannon Index, BMI and sex) and multiple testing.

(2) The VFM diet score and 15 food groups not forming the score were fitted into a backward stepwise linear regression model to predict each significant taxon using *P*<0.05 as the cut off threshold.

(3) All 20 food groups were fitted into a backward stepwise linear regression model to predict each significant taxon using *P*<0.05 as the cut off threshold. Significant results shown only for foods forming the VFM diet score.

(4) *Eubacterium dolichum* is the only taxon associated with VFM independently of the VFM diet score (Beta[SE]: 0.057[0.019], $P=2.74x10^{-3}$).

7.4 Discussion

In this chapter, I developed a dietary VFM risk score and evaluated it against blood metabolomics and faecal microbiome profiles. I characterised the metabolomics profile of the VFM dietary pattern and additionally found a specific gut bacterial species to associate with this pattern and VFM independently of confounders such as age, BMI and total fat mass. In particular, using this novel dataset I found the species *E. dolichum* in the gut and hippurate in blood may connect an unhealthy diet to VFM.

The score I created was strongly predictive of VFM in the TwinsUK cohort, and it was replicated VFM-discordant MZ twins. This allowed me to further explore the potential effects of diet on VFM development through metabolomics and microbiome methods. Overall, I found 4 metabolites were significantly associated with both the VFM diet score and VFM (after adjusting for diet). These metabolites also accounted for some portion of the variance in the association between the VFM diet score and VFM. Specifically lower levels of hippurate and bilirubin (Z,Z), and higher levels of alpha-hydroxyisovalerate and butyrylcarnitine were associated with higher VFM diet scores and VFM. Both alpha-hydroxyisovalerate and butyrylcarnitine are metabolites involved in BCAA catabolism and fatty acid metabolism, both have previously been shown to be higher in obese children (Butte et al., 2015) and adults (Moore et al., 2014). In particular, alpha-hydroxyisovalerate has been found to predict insulin resistance and glucose intolerance well (Gall et al., 2010, Varvel et al., 2014). Higher levels of both alpha-hydroxyisovalerate and butyrylcarnitine were associated with increased intakes of red and processed meats and eggs. Although animal sources of fats and protein have yet to be connected to deranged BCAA metabolism in humans, a previous study has shown through feeding mice a HF diet the addition of BCAA (which are primarily sourced from animal-derived proteins) aggravates insulin resistance by activating the mTOR kinase pathway (Newgard et al., 2009).

Bilirubin is an endogenous anti-oxidant that also has roles in haemoglobin and porphyrin metabolism. Similar to my findings, reduced levels of serum bilirubin have been associated with increased abdominal adiposity and metabolic derangements (Wu et al., 2011b, Kwon et al., 2011, Jenko-Praznikar et al., 2013). I found increased reported consumption of fried and fast foods were associated with reduced bilirubin (Z,Z). A previous study identified reported intakes of total fatty acids were correlated with reduced bilirubin in serum (Jenko-Praznikar et al., 2013). This association may be potentially derived from elevated levels of

oxidative stress depleting bilirubin. In support of this, increased intakes of fried and fast foods were associated with reduced levels of EPA and increased proline (related to collagen breakdown (Palka, 1996)) that might suggest higher inflammation and oxidative stress (Krishnan et al., 2008). Moreover, vegetable oil frying also lowers polyphenols and was shown to increase liver microsomal lipid peroxides when fed to mice (Quiles et al., 2002).

The primary metabolite connecting diet to VFM was hippurate. Overall, reduced levels of hippurate in blood were associated with higher VFM, a finding that I additionally replicated in VFM-discordant MZ twins, the VFM dietary score and *E. dolichum* in the gut. Hippurate is a glycine conjugate of benzoic acid formed in the mitochondria of the liver (Gatley and Sherratt, 1977) and kidneys (Temellini et al., 1993), but is also a is a mammalian-microbial co-metabolite formed by gut bacterial metabolism of dietary components, in particular polyphenols (Gonthier et al., 2003, Walsh et al., 2007), to produce benzoic acid. I found increased levels of hippurate to be independently associated with higher intakes of fruit and whole grain products, which are high in polyphenols. Concerning obesity and metabolic diseases animal models which have shown reduced hippuric acid urinary excretion in obesity (Shearer et al., 2008, Waldram et al., 2009, Calvani et al., 2010) and higher levels in Type II diabetes (Williams et al., 2005) compared to controls. To my knowledge no human studies to date have extensively evaluated the relationship between VFM and blood metabolite biomarkers of dietary polyphenol consumption, including hippurate. Although in the previous chapter I found hippurate to be positively associated with microbiome diversity as well as inversely associated with MetS risk. Interestingly, I also identified higher levels of hippurate in blood to be associated with increased expression of neuroglobin in adipose tissue. Neuroglobin, is a globin protein expressed mainly in neurons and some endocrine tissues (Burmester et al., 2000), it has roles in protecting cells against hypoxia and oxidative stress (Burmester et al., 2007). Hypoxia, which is a lack of oxygen reaching tissues, has emerged in recent studies as an important mechanism in the progression of adipose tissue dysfunction (Kim et al., 2014). This finding suggests a mechanism in which hippurate could protect against adipose tissue dysfunction and resultant VFM development.

E. dolichum is a bacterial species belonging to the family *Erysipelotrichaceae*, increased abundances of *E. dolichum* were associated with higher dietary VFM scores and VFM (independently of diet). This finding suggests this microbe may contribute to VFM development through diet, notably lower intakes of whole grains. Similar to my findings, in a

murine model of Western-style diet induced obesity, abundances of *E. dolichum* were increased (Turnbaugh et al., 2008), additionally further metagenomics analysis showed the *E. dolichum* genome may be enriched for genes that allow this microbe to have an advantage under a Western-style diet as it is enriched for phosphotransferase proteins which have roles in the import and metabolism of simple sugars. An additional study of Japanese quail strains (atherosclerotic-resistant and non-resistant) showed increased *E. dolichum* abundance in atherosclerotic-resistant quails compared to control when fed a high cholesterol diet (Liu et al., 2015). I believe this is the first human study to find a connection between *E. dolichum* and a diet high in fat and low in fibre and micronutrients however I do not believe there to be literature on the mechanisms of the effect of this species on VFM development. It may be that the relationship between *E. dolichum* and VFM is primarily an artefact of poor diet and not specifically contributing to VFM, however the association was unaffected when adjusted for the diet score. Within TwinsUK higher abundances of *E. dolichum* have associated with increased scores on the frailty index, a finding that remained strong after adjustment for diet and diversity (Jackson et al., 2016b). The association I found between *E. dolichum* and hippurate is difficult to untangle with my limited dataset, though should be assessed in future studies.

This chapter had many strengths. I believe this to be the first large-scale study applying multi-omic methods to explore the effect of diet on VFM. The food intake associations with VFM replicate findings from previous studies, this confirms the strength and validity of the dietary VFM score. In particular, I replicated associations from previous studies finding higher reported intakes of fruit (Romaguera et al., 2011), dairy products (Romaguera et al., 2011), and whole grains (Caron-Jobin et al., 2011) with reduced VFM or WC, while higher intakes of fried and fatty foods (Mollard et al., 2014, Romaguera et al., 2011) and red and processed meats (Romaguera et al., 2011) with increased VFM or WC. Similar to VFM (Direk et al., 2013), I found the dietary risk score determined strongly by genetics (h^2: 44%) this is in line with previous studies suggesting the heritability of an 'unhealthy' diet pattern ranges from 33 to 50 % (Pallister et al., 2014). I also experienced a number of limitations throughout this chapter. The study population was mainly females and my results may not apply to men. Due to the cross-sectional nature of the study, I cannot attribute cause and effect to my findings. I adjusted for a large number of potential confounders, though I still face possible residual or unmeasured confounding due to other unmeasured factors or measurement error. Although, as I did adjust for a large set of confounders as well as for multiple testing it is not likely that these factors

would totally account for my results. The measurement of the gut microbiome was further limited by using 16S gene sequencing. Subsequent investigations using more accurate metagenomic methods may allow for a more thorough examination into microbe-metabolite interactions at the functional level. I used a variety of time points for sampling, however it is likely my results may be enhanced if the same time point was used throughout. I also did not have repeated measurements for individuals and therefore I could not evaluate intra-individual variation. Finally, I was not able to replicate my findings in an independent population, despite this I was able to replicate top association in MZ VFM-discordant twins that are matched for age, gender and the baseline genetic sequence.

Conclusions

An unhealthy dietary is an important influence on VFM. In this chapter I connected a dietary VFM score and VFM to a single gut microbial species and blood metabolites. In particular, I found the species *E. dolichum* seems to link a dietary pattern with low in fruit, whole grains and fermented dairy product intakes and high red, processed meat and eggs, and fried and fast foods intakes to VFM, which was independent of BMI and total fat mass. Additionally, I found hippurate, an important benzoate and microbial co-metabolite, to link these features to the microbiome. Finally, I found hippurate to associate with neuroglobin expression in adipose tissue, alluding to a potential mechanism of interaction. In the future, studies may aim to verify my findings through dietary intervention and establish further how our findings influence long-term metabolism and health.

Chapter 8 Conclusions and suggested future directions

Metabolomics has emerged from this work as a key application for furthering nutrition research. Through examining the metabolome in a bodily system, specific organ, tissue or cell, information is provided as to the metabolic state of an individual. Nutrition and the metabolome are highly related as nutrients and non-nutrient food constituents supply chemical components to the body, although separating these contributions can be challenging (Gibney et al., 2005). In spite of this, using self-reported food intakes candidate biomarkers have been identified within the metabolome in this study and others (O'Sullivan et al., 2011, Altmaier et al., 2011, Guertin et al., 2014b), validating this approach. However, the genetic effect on the metabolite profile is highly variable (Suhre et al., 2011), as such application of the co-twin control method can provide an excelled way to further metabolomics studies by providing cases and controls matched for age, sex and genetics.

In this thesis, through using metabolomics in collaboration with self-reported dietary intakes I have identified a number of candidate biomarkers that can be used and assessed in subsequent epidemiology studies. I evaluated the utility of using metabolite scores to assess food intakes and applied the scores to study the odds of developing MetS. I used metabolites to investigate the metabolomics signature of gut microbiome diversity and connected increased levels of hippurate in blood to intakes of foods high in polyphenols, diversity and lower odds of developing MetS. Finally, I attempted to untangle the influence of a diet predictive of VFM on VFM development and found blood levels of hippurate and *E. dolichum* in the gut may be important links. Throughout each chapter I used discordant MZ twins who are controlled for age, sex and the baseline genetic sequence to validate my findings. In the discussion below I have highlighted the key findings from each chapter.

Key points/learning from the thesis

Chapter 4: Metabolite-food intake discovery analysis

- Self reported food intakes correlate strongly with levels of blood metabolites.
- Of 601 total metabolites assessed, 128 metabolites (21.3%) were associated with food intakes and dietary patterns using stringent cut-offs for multiple testing.
- Wine intake and seafood/fish intakes yielded the strongest and most abundant results, many replicated from previous studies.

- 72 food intake-metabolite associations were novel: including associations between mushroom intake and ergothioneine, apples/pear and 3-phenylpropionate, and meat intake and trans-4-hydroxyproline.
- Assessing both dietary patterns and specific food group intakes yielded unique results.

Chapter 5: Metabolite-food scores

- Creating metabolite scores provided an advantage over using single metabolites to measure food intakes.
- Differences between each of the methods to create the scores were negligible, though the weighted method performed the best.
- Metabolite scores performed moderately for alcohol (AUC>0.8) and (AUC>0.7) for fruit, whole grains, seafood, fried foods, tea and coffee.
- Metabolite scores for seafood, whole grains, and butter and creams were associated with a reduced odds of future MetS, a finding not replicated by intake data.

Chapter 6: Metabolite markers of gut microbiome diversity

- Gut microbiome alpha-diversity was associated with five blood metabolites: hippurate, p-cresol sulfate, phenylacetylglutamine and 3-phenylpropionate which were associated positively; and negatively with hyodeoxycholate
- Hippurate, a benzoate metabolite derived from polyphenols, was associated with intakes of fruit and whole grains cross-sectionally and longitudinally.
- Hippurate was positively associated with OTUs and collapsed taxa within the family Ruminococcacea.
- Hippurate was negatively associated with OTUs and collapsed taxa within the Firmicutes phyla, the family Lachnospiraceae, the class Erysipelotrichi, and the order Actinomycetales.
- Longitudinal hippurate levels were associated with a reduced odds of MetS, which was largely mediated by Shannon diversity (61.1%).

Chapter 7: Dietary VFM score and metabolomic and microbiome profiles

- Creating a diet score was a robust means to predict VFM
- The dietary risk score had a strong metabolomics signature: associating with 30 metabolites
- Notably, in blood lower hippurate and bilirubin (Z,Z) levels, and higher alpha-hydroxyisovalerate and butyrlcarnitine levels were associated to increased VFM and the VFM diet score.
- Increased abundances of the collapsed taxonomy assigned to *Eubacterium dolichum* was associated with both increased VFM and the VFM diet score
- Increased *Eubacterium dolichum* adbundances were associated with reduced blood hippurate and adipose tissue expression of neuroglobin, a protein protective of hypoxia and oxidative stress.

Chapter 4: Metabolite-food intake discovery analysis

I explored all metabolomic associations with reported food intakes and dietary patterns and used MZ twins discordant for food intakes to validate the findings. Overall, I identified and validated 128 metabolites to be associated between self-reported food intakes and diet patterns. To my knowledge, 72 metabolite associations with food intakes were novel (have not been identified in blood previously by studies also using FFQ data (Guertin et al., 2014b, Zheng

et al., 2014)). Identifying this many novel associations is encouraging for future dietary metabolomics studies suggesting there is much potential within the metabolome. Moreover, that I replicated many findings from previous studies further supports the methods used. Interestingly, 21 metabolites were associated to dietary patterns not items which may suggest that through measuring dietary patterns we are able to capture information not obtained by measuring single foods independently. In my analysis, I did not attempt to assess the utility of using single food items versus combined biomarkers of dietary patterns when applied to other indicators of health such as MetS, fasting blood glucose, waist circumference or BMI, which could be an area for future studies. Certainly a larger number of metabolites representing diverse metabolic pathways are associated with dietary patterns highlighting this potential area of importance. Although, my results with principal components patterns are population-specific and therefore cannot be replicated, which is why I chose to analyse the Mediterranean diet using an *a priori* method. This method yielded a lower number of associations than the comparable Fruit and Vegetable PC pattern, though this also highlights that the Mediterranean diet score has limitations, which have been outlined previously (Hoffman and Gerber, 2013). Moreover, I did not find any metabolite-associated SNPs to be related to food intake, though more precise food intake data or using combined genetic risk may have a stronger effect. Moreover, testing these SNPs or combined genetic risk SNPs to examine gene-diet interactions are a potential area for future study. I used discordant MZ twins as a validation population which was very stringent criteria, though this strongly supports the value of the associations I identified. Future large-scale feeding studies need to be undertaken to confirm our associations, to properly quantify effects and also to determine the origin of the associations (e.g. whether the metabolite is derived from a food itself and its quantity in food). Further adding to the complexity of associations are intermediate processes such as one's genetic makeup and the influence of gut bacteria metabolism. Finally, 32 metabolites that were associated to food intakes and 29 to dietary patterns are currently not identified but in the future may provide useful metabolic insights.

Chapter 5: Metabolite-food scores

For the Chapter 5 analysis, I created metabolite scores (when there were multiple metabolites associated to a food) using three methods (quartile ranking, continuous scoring and weighted) and compared each method against the top associated metabolite. I believe this was the first

time this was done before and certainly with such a large dataset. In the training dataset, I was able to form metabolite scores for fifteen different food groups (vegetables, fruit, whole grains, nuts and legumes, seafood, white meat, red meat, fried foods, sweets and sweet baked products, butter and cream, spreads and dressings, milk, tea, coffee and alcohol). To form these groups I used highly stringent methods: associations must have passed multiple testing cut-offs, been replicated in MZ discordant twins and maintained significance following the data reduction steps. Of course, the methods I used may have been too stringent and I may have formed more powerful groups if I had used more metabolites, on the other hand, if my results were to be replicated this would be easier done with fewer metabolites. Nonetheless, I replicated my results in the testing group, therefore at least some of these scores should translate as important markers in independent populations. As for the performance of the scores, they generally performed better than the top single metabolite and the weighted method tended to perform best although differences were negligible. At predicting high and low consumers in the test group, metabolite scores performed well (AUC>0.8) for alcohol and reasonably (AUC>0.7) for fruit, whole grains, seafood, fried foods, tea and coffee. Moreover, I applied a unique method by using food group liking as an alternative way of indicating food intakes, I found metabolite scores for all groups but vegetables, white meat and spreads and dressings to correlate with reported food liking which also supported the results of the ROC analysis. Based on my findings from this analysis, each scoring method appears to be suitable for use in biomarker studies, though due to its simplicity, the continuous method (multiplying metabolite concentrations according to the association direction and summing) may be ideal, especially in studies using multiple populations.

In the second part of the chapter, I used validated metabolite scores to examine their association with MetS and its components in the twins at least 5 years after metabolomics profiling. I found high levels of metabolites correlated to consumption of seafood, whole grains and butter and cream intakes to associate with a lower risk of MetS. Although reported intakes did not reflect these findings in our sample. However, the dataset I used to examine these associations was very small and only 25 twins developed MetS in later years. Therefore, further analysis could be done on dataset with a larger number of cases. Although, the ability to identify an association with MetS with such a small sample size might be encouraging for the usage of metabolomics for future studies. Also, the potential applicability for these metabolite biomarkers

to be used as pre-disease markers of dietary risk in a personalised medicine setting is an exciting area for further development.

Chapter 6: Metabolite markers of gut microbiome diversity

I found five blood metabolites correlated with gut microbiome diversity (hippurate, p-cresol sulfate, phenylacetylglutamine and 3-phenylpropionate which were associated positively; and negatively with hyodeoxycholate) and examined their association with diet and relationship with MetS. **Figure 8-1** shows a summary of the findings. Overall, higher levels of hippurate in blood were the strongest correlate of increased gut microbiome diversity, which was validated in MZ twins discordant for diversity. Hippurate was also associated with reported food intakes and I therefore examined it more thoroughly. Hippurate is a metabolite derived from benzoic acid which is primarily produced by gut bacterial metabolism of dietary phenolic acid (Gonthier et al., 2003). Supported by previous studies (Gonthier et al., 2003, Walsh et al., 2007), reported intakes of high polyphenol-containing foods, whole grains, fruits and coffee, were significantly predictive of longitudinal changes in blood hippurate levels. In the faecal microbiome I characterised a gut microbiome profile correlated with blood hippurate levels and overall identified 46 associations. This included higher hippurate levels being associated with lower levels of Firmicutes phyla overall, and lower and higher levels of Lachnospiraceae and Ruminococcaceae, respectively. Interestingly, both Lachnospiraceae and Ruminococcaceae are butyrate-producing bacteria (Vital et al., 2014), therefore measuring faecal short-chain fatty acid production may be an interesting area future investigation. Comparing those OTUs/taxa associated with hippurate, I found 5 OTUs and 5 collapsed taxa to be associated to the the hippurate diet score as well, suggesting food intake may modify hippurate production through enriching these bacteria, though this would need to be confirmed by feeding studies. I had the advantage of using longitudinal hippurate data and found an increasing trajectory to be associated with a reduced odds of later having MetS, reduced BMI and TG. The longitudinal metabolite data certainly provided a strong advantage to this analysis, though Shannon diversity accounted for 61.1% of the variance in the association between the hippurate trajectory and MetS, suggesting that Shannon diversity was strongly driving this association. Although, the dietary factors were not associated with MetS, which may be related to less measurement precision. Untangling the dietary influence of hippurate and Shannon diversity on MetS and its components is challenging due to the nature of FFQ data. It is likely that food intakes are

influencing diversity, the gut microbiome profile and hippurate production though separating the effect of diet is a challenge. It is important to consider whether there are diet-microbe interactions in the formation of hippurate, I attempted to investigate this but did not find any associations, potentially due to the limited precision of the dataset but also the time between sample collection and FFQ completion may be of relevance. It is likely some individuals have a gut microbiome profile with an increased capacity for hippurate production and therefore increased ability to handle oxidative stress, a central component of metabolic diseases. However no clear patterns of microbes emerged. Future dietary intervention studies should test our findings by feeding individuals high polyphenol-containing foods and testing the changes induced on the microbiome, and blood hippurate levels. Longer-term studies should explore the influence of hippurate on MetS.

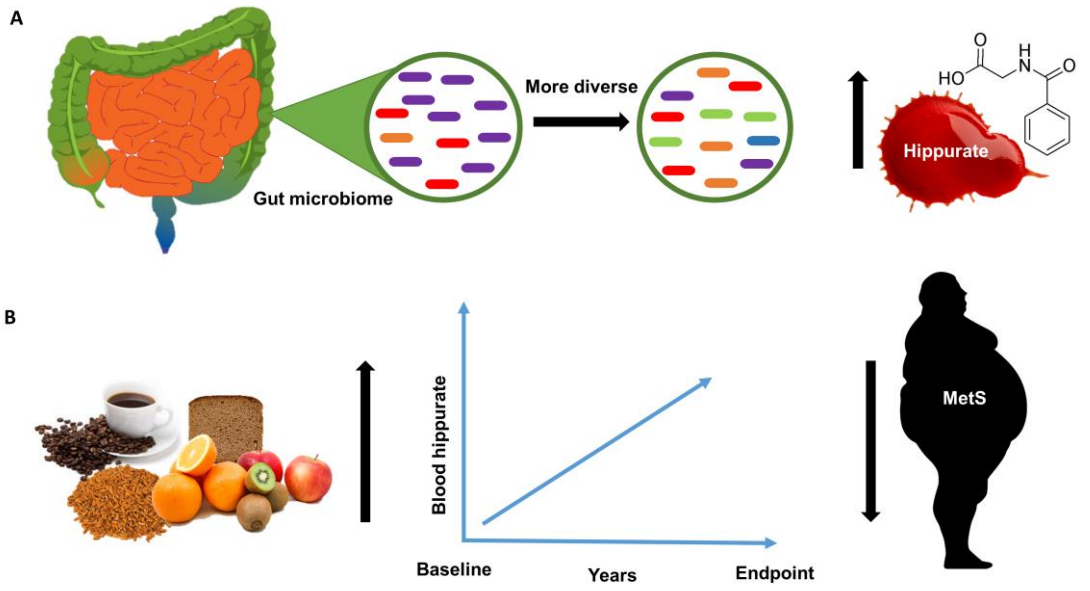


Figure 8-1. Summary of findings from Chapater 6
Figure 8-1a represents increased gut microbiome diversity correlates with increased blood hippurate. Figure 8-1b represents increased intakes of high polyphenol foods (whole grains, coffee and fruits) predict an increasing hippurate trajectory. Moreover, an increasing hippurate trajectory correlates with lower risk of MetS (metabolic syndrome) at endpoint. Images have been obtained and modified from google search under the usage right option 'labelled for reuse with modification'.

Chapter 7: Dietary VFM score and metabolomic and microbiome profiles

I identified foods most strongly associated with VFM and developed and validated a predictive dietary VFM-risk score. I also characterised the blood metabolomics and gut microbiome profiles associated with the dietary VFM-risk score and VFM. **Figure 8-2** shows a summary of the findings. I found reported intakes of fried and fast foods and red, processed meats and eggs

were associated with increased VFM while intakes of fruits, fermented dairy products, and whole grain products were associated with lower VFM, moreover the dietary risk score was highly predictive of VFM correctly identifying 92.7% of twins in the upper and lower tertiles of VFM. The dietary risk score had a strong metabolomics signature, associating with 30 metabolites, many of which were associated to food intakes in the first and second chapters. After data reduction, I found 4 metabolites were associated with both increased VFM and the VFM-diet, including reduced hippurate and bilirubin (Z,Z) (involved in hemoglobin and porphyrin metabolism), and increased alpha-hydroxyisovalerate (involved in valine, leucine and isoleucine metabolism) and butyrylcarnitine (involved in fatty acid and BCAA metabolism). It is encouraging that hippurate was found to associate with future MetS in Chapter 6 and also with VFM, as waist circumference is a component of MetS.

The VFM diet score was also associated with 6 collapsed taxa and 8 OTUs, within the faecal microbiome, though only 1 *Bifidobacterium* OTU and 1 taxa, *Eubacterium dolichum* was associated with both VFM and the VFM-diet. Interestingly, increased levels of *E. dolichum* associated with a reduced level of hippurate in blood. Upon further investigation I found increased hippurate in blood to be associated with elevated adipose tissue expression of neuroglobin (Burmester et al., 2000) a protein which protects cells against hypoxia and oxidative stress (Burmester et al., 2007), the importance of hypoxia in the development adipose tissue dysfunction has recently emerged (Kim et al., 2014). This may also suggest that hippurate protects against adipose tissue dysfunction and VFM development as a result. It would be useful if mechanistic studies were performed to confirm if hippurate plays a protective role in adipose tissue dysfunction and also, importantly, determine the function of neuroglobin in adipose tissue as no studies have outlined this before. Abundance of the species *Eubacterium dolichum* has previously shown to increase in mice fed a Western-style diet high in saturated fat and simple carbohydrates and low in plant polysaccharides (Turnbaugh et al., 2008). This study on mice found *E. dolichum* to function in the import and processing of simple sugars (Turnbaugh et al., 2008), using metagenomics analysis of the microbiome data from my study could be used to confirm their findings (Turnbaugh et al., 2008). To my knowledge this is the first time a relationship between *Eubacterium dolichum* and a poor diet has been made in humans and the usefulness of this marker for epidemiology and clinical studies when diet histories are difficult to obtain should be evaluated.

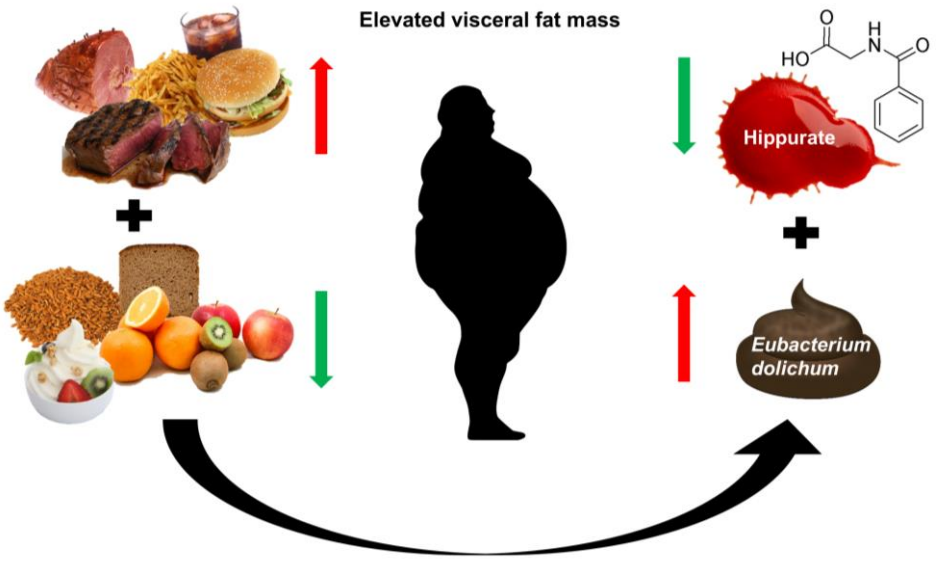


Figure 8-2. Summary of findings from Chapter 7
Figure 8-2 represents the visceral fat mass (VFM) diet which is composed of high intakes of fried and fast foods and red processed meat and eggs and low intakes of fruits, whole grains and fermented dairy products. The VFM diet and VFM correlate with lower levels of blood hipurate and faecal *Eubacterium dolichum*. The color of the arrows represent direction protective of VFM: *red*, increased VFM; *green*, lower VFM. Images have been obtained and modified from google search under the usage right option 'labelled for reuse with modification'.

What I learned and challenges

In general I found using FFQ data throughout the thesis to be problematic. Though I did have success in finding associations, to bring the field forward we will need to step away from using FFQs and begin to apply intervention studies or use more detailed dietary data (such as food records). I also found defining food intake groups to be problematic. Although it is a common practice in nutritional epidemiology, statistically intakes of the foods included within these food groups are not always correlated and therefore I found providing a logical reasoning for group assignment or understanding the biological mechanisms for the metabolite associations was challenging. Moreover, the relative importance of a single food as opposed to dietary patterns is difficult to ascertain. I did learn how to conduct a principal component dietary pattern analysis, though whether these patterns are more reflective of patterns in reporting rather than true behaviours is a question that requires asking. I recognise that FFQ data is the best data I had access to for the study and for some foods it is better at capturing habitual consumption for (e.g. coffee and alcohol).

I also learned that there are many ways to analyse metabolomics data. In the field of metabolomics it is a common practice to perform dimension reduction techniques (such as PCA or partial least squares regression) due to the correlation between metabolites. The results of

my thesis may have been quite different if I had first undertaken dimension reduction techniques. For one, I would have had less risk for false negative associations with fewer variables. However, it is likely that not all of the metabolites included in latent variables (e.g. PCs) would be directly related to the intake of a food, similar to the issue that not all foods included in a food group are likely to explain a metabolite association. Had I undertaken dimension reduction it would also have been challenging to unravel the reasoning for correlations between certain metabolites.

Throughout my work I learned and employed the co-twin control method that was a unique aspect of the thesis as a whole. One issue I did have was with defining discordance for food intake data. As FFQ data is not precise using a 1 SD cut-off for discordance may not have been an ideal definition for all foods. The significant food-metabolite associations I found in discordant MZ twins may have just been a reflection of improved accuracy of the FFQ to measure certain food intakes. It may have been better to define MZ discordance as one twin consuming a food while the other does not, though likely there would have been fewer instances of these and it would have been food-dependent. The twin model does have potential for further exploitation especially in the context of dietary intervention studies, where twin studies are few. Despite all of these challenges I encountered, I do believe that metabolomics has strong potential for improving our understanding of the relationship between diet and disease and we should continue the work being done in this field.

Future directions and conclusion

Dietary metabolomics research is still very much in its infancy and there are a number of issues which need to be overcome before the ideal clinical test panel of dietary biomarkers to measure food intake status can be achieved. To take the field forward, I have specific recommendations regarding the work I undertook in this thesis (providing Chapter 7 as an example) and general recommendations concerning the field as a whole.

With regards to carrying the work I undertook in this thesis forward, I would recommend the following:

1. Replication of significant results in independent cohorts using similar datasets
2. Repeating the analysis with new metabolomics platforms
3. Dietary intervention:
    a. To confirm and quantify the food intake effect on biomarkers

b. To evaluate metabolomics and microbiome findings with the metabolic phenotypes studied (VFM and MetS [plus components])

Due to the very good predictive performance of the VFM dietary risk score, I would suggest undertaking a 4- or 5-week dietary intervention modelled after the score. Ideally, this could also be done as a cross-over study where for one phase a VFM-risk diet is consumed while during the other phase a VFM-protective diet is consumed, separated by a washout period. Therefore, for the VMF-risk phase, the diet would consist of low intakes of fruit, whole grains and fermented dairy products and increased intakes of fried and fast foods and red, processed meat and eggs, and the opposite for the VFM-protective phase. A challenge may be determining the ideal quantities of foods to be consumed and the incorporation of other foods to ensure a balanced diet. This could involve collecting subject food records prior to beginning the study and personalising the diets to ensure compliance and real-world applicability of the findings, though this would utilize more time and resources. Visceral fat mass, faecal microbiome and blood metabolomics data would be measured at the beginning and end of each phase.

To analyse the data I would aim to confirm the findings from my analysis in the greater twin dataset. Therefore I hypothesize that following a diet high in fried and fast foods and red, processed met and eggs, and low in fruit, whole grains and fermented dairy products will increase VFM, alpha-hydroxyisoverate, butyrlcarnitine and *E. dolichum*, and reduce levels of hippurate and bilirubin (Z,Z). However, if the same technologies were used it could be interesting to confirm if overall results using all metabolites and microbiome phenotypes are comparable to my findings in the larger twin dataset. Moreover, twin modelling could be an asset, for instance to determine if MZ twins are more alike than DZ twins in their response to the intervention, or through using the co-twin control method, which in this case could be confirming the hypothesis using MZ twins discordant for feeding (e.g. comparing endpoint phenotypes of one twin fed the VFM-risk diet compared to baseline phenotypes of other twin).

General areas to further nutritional metabolomics studies

For the future of nutritional metabolomics there are a number of issues to be resolved, some of which I encountered, and areas of research which require further exploration to strengthen the field. For instance, many of the associations I identified were with metabolites endogenous to the human body and therefore determining the origin of these metabolites and quantifying the

effect of food consumption on levels is complex. The food metabolome potentially contains many metabolites that are exogenous to the human body and therefore are less influenced by metabolic processes, though it may be more difficult to obtain biological information on these metabolites. Delving deeper into the food metabolome therefore may be useful, this would also involve compiling more complete metabolomics profiles of foods into databases, which is now already taking place (Scalbert et al., 2014), such as the FooDB (www.foodb.ca), Phenol-Explorer (www.phenol-explorer.eu) and PhytoHub (www.phytohub.eu). Moreover, careful intervention studies are needed to evaluate the utility of biomarkers, whether they are applicable for short (hours/days) or long-term usage (months/years), as well as the relevance of the type of biosample provided (e.g. urine vs blood). Feeding studies evaluating short-term biomarkers within exposure are of course the easiest to undertake whereas longer-term studies will be more challenging due to difficulties such as participant compliance and cost factors, however these will be more valuable.

The generalisability of my findings to both males and females and determining racial effects have also not been thoroughly explored here nor in many other studies. One study found one third of metabolites were influenced by sex, with network analysis revealing sex-specific metabolic pathways (Krumsiek et al., 2015). Racial effects have not been well characterised, though one study found distinct differences between the metabolomics profiles of Caucasians versus African Americans in response to atenolol treatment (Wikoff et al., 2013), I did however confirm some of the top findings from two food metabolomics studies on African Americans (Zheng et al., 2014).

Longitudinal data of metabolite levels, which I was able to have access to for hippurate, will likely provide stronger correlations with intakes and power for outcome prediction, improving studies more still. In additions, repeating the findings of my study with more advanced platforms measuring more metabolites and with increased quantitative accuracy (such as NMR-based metabolomics) will likely generate further novel findings. Applications of food intake biomarkers to study disease outcomes is one of the primary aims of the research and future studies should begin to apply these methods, for which longitudinal data will be a particular asset.

Interactions between the human gut microbiome and diet and the subsequent influence on metabolism are not well characterised at this time therefore future studies are needed. One way this could be taken further is through re-analysing the gut microbiome data used in my study using metagenomics data to determine the metabolic functions of the bacteria associated

with hippurate or the VFM diet. Analysing the faecal metabolome would add further information. Another continuing issue with microbiome data is that methods have not been effectively established, for instance studies are quite heterogeneous with regards to quality control procedures and referencing panels used making cross-study comparisons difficult. Even within my thesis, two different referencing methods were used for chapters 6 and 7. This meant that it was difficult to compare results across the thesis, it may have been useful if I had repeated the analysis using the same datasets for both chapters. To aid this and other issues collaboration between cohorts is essential.

Conclusion

In conclusion, I have shown that metabolomics has great promise for furthering nutrition research. Using high-throughput omic data and a unique human cohort I have generated a number of novel and exciting findings. In particular, I found unique correlations between food intakes and metabolites, I evaluated different methods for creating combined metabolite scores, I identified hippurate as a useful blood marker of gut microbiome diversity that is also related to diet and MetS and VFM. I also characterised the metabolomics and microbiome profile of an unhealthy diet associated with VFM. I have shown the broad applications of metabolomics to nutritional epidemiology which should stimulate further work in this area.

References

ABDULLAH, M. M., CYR, A., LEPINE, M. C., LABONTE, M. E., COUTURE, P., JONES, P. J. & LAMARCHE, B. 2015. Recommended dairy product intake modulates circulating fatty acid profile in healthy adults: a multi-centre cross-over study. *Br J Nutr,* 113**,** 435-44.

ADEN, D. P., FOGEL, A., PLOTKIN, S., DAMJANOV, I. & KNOWLES, B. B. 1979. Controlled synthesis of HBsAg in a differentiated human liver carcinoma-derived cell line. *Nature,* 282**,** 615-6.

AGRAWAL, A. & LYNSKEY, M. T. 2008. Are there genetic influences on addiction: evidence from family, adoption and twin studies. *Addiction,* 103**,** 1069-81.

ALBERTI, K. G., ECKEL, R. H., GRUNDY, S. M., ZIMMET, P. Z., CLEEMAN, J. I., DONATO, K. A., FRUCHART, J. C., JAMES, W. P., LORIA, C. M. & SMITH, S. C., JR. 2009. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation,* 120**,** 1640-5.

ALCOCK, J., MALEY, C. C. & AKTIPIS, C. A. 2014. Is eating behavior manipulated by the gastrointestinal microbiota? Evolutionary pressures and potential mechanisms. *Bioessays,* 36**,** 940-9.

ALEXANDER, D., LOMBARDI, R., RODRIGUEZ, G., MITCHELL, M. M. & MARIAN, A. J. 2011. Metabolomic distinction and insights into the pathogenesis of human primary dilated cardiomyopathy. *Eur J Clin Invest,* 41**,** 527-38.

ALTMAIER, E., KASTENMULLER, G., ROMISCH-MARGL, W., THORAND, B., WEINBERGER, K. M., ILLIG, T., ADAMSKI, J., DORING, A. & SUHRE, K. 2011. Questionnaire-based self-reported nutrition habits associate with serum metabolism as revealed by quantitative targeted metabolomics. *Eur J Epidemiol,* 26**,** 145-56.

ANHE, F. F., ROY, D., PILON, G., DUDONNE, S., MATAMOROS, S., VARIN, T. V., GAROFALO, C., MOINE, Q., DESJARDINS, Y., LEVY, E. & MARETTE, A. 2015. A polyphenol-rich cranberry extract protects from diet-induced obesity, insulin resistance and intestinal inflammation in association with increased Akkermansia spp. population in the gut microbiota of mice. *Gut,* 64**,** 872-83.

ANZMAN-FRASCA, S., SAVAGE, J. S., MARINI, M. E., FISHER, J. O. & BIRCH, L. L. 2012. Repeated exposure and associative conditioning promote preschool children's liking of vegetables. *Appetite,* 58**,** 543-53.

ARUMUGAM, M., RAES, J., PELLETIER, E., LE PASLIER, D., YAMADA, T., MENDE, D. R., FERNANDES, G. R., TAP, J., BRULS, T., BATTO, J. M., BERTALAN, M., BORRUEL, N., CASELLAS, F., FERNANDEZ, L., GAUTIER, L., HANSEN, T., HATTORI, M., HAYASHI, T., KLEEREBEZEM, M., KUROKAWA, K., LECLERC, M., LEVENEZ, F., MANICHANH, C., NIELSEN, H. B., NIELSEN, T., PONS, N., POULAIN, J., QIN, J., SICHERITZ-PONTEN, T., TIMS, S., TORRENTS, D., UGARTE, E., ZOETENDAL, E. G., WANG, J., GUARNER, F., PEDERSEN, O., DE VOS, W. M., BRUNAK, S., DORE, J., ANTOLIN, M., ARTIGUENAVE, F., BLOTTIERE, H. M., ALMEIDA, M., BRECHOT, C., CARA, C., CHERVAUX, C., CULTRONE, A., DELORME, C., DENARIAZ, G., DERVYN, R., FOERSTNER, K. U., FRISS, C., VAN DE GUCHTE, M., GUEDON, E., HAIMET, F., HUBER, W., VAN HYLCKAMA-VLIEG, J., JAMET, A., JUSTE, C., KACI, G., KNOL, J., LAKHDARI, O., LAYEC, S., LE ROUX, K., MAGUIN, E., MERIEUX, A., MELO MINARDI, R., M'RINI, C., MULLER, J., OOZEER, R., PARKHILL, J., RENAULT, P., RESCIGNO, M., SANCHEZ, N., SUNAGAWA, S., TORREJON, A., TURNER, K., VANDEMEULEBROUCK, G., VARELA, E., WINOGRADSKY, Y., ZELLER, G., WEISSENBACH, J., EHRLICH, S. D. & BORK, P. 2011. Enterotypes of the human gut microbiome. *Nature,* 473**,** 174-80.

BALLATORI, N., FANG, F., CHRISTIAN, W. V., LI, N. & HAMMOND, C. L. 2008. Ostalpha-Ostbeta is required for bile acid and conjugated steroid disposition in the intestine, kidney, and liver. *Am J Physiol Gastrointest Liver Physiol,* 295**,** G179-g186.

BARAWKAR, D. A., BANDYOPADHYAY, A., DESHPANDE, A., KOUL, S., KANDALKAR, S., PATIL, P., KHOSE, G., VYAS, S., MONE, M., BHOSALE, S., SINGH, U., DE, S., MERU, A., GUNDU, J., CHUGH, A., PALLE, V. P., MOOKHTIAR, K. A., VACCA, J. P., CHAKRAVARTY, P. K., NARGUND, R. P., WRIGHT, S. D., ROY, S., GRAZIANO, M. P., CULLY, D., CAI, T. Q. & SINGH, S. B. 2012. Discovery of pyrazole carboxylic acids as potent inhibitors of rat long chain L-2-hydroxy acid oxidase. *Bioorg Med Chem Lett,* 22**,** 4341-7.

BARRIOS, C., BEAUMONT, M., PALLISTER, T., VILLAR, J., GOODRICH, J. K., CLARK, A., PASCUAL, J., LEY, R. E., SPECTOR, T. D., BELL, J. T. & MENNI, C. 2015. Gut-Microbiota-Metabolite Axis in Early Renal Function Decline. *PLoS One,* 10**,** e0134311.
BEASLEY, T. M., ERICKSON, S. & ALLISON, D. B. 2009. Rank-based inverse normal transformations are increasingly used, but are they merited? *Behav Genet,* 39**,** 580-95.
BENE, J., KOMLOSI, K., GASZTONYI, B., JUHASZ, M., TULASSAY, Z. & MELEGH, B. 2005. Plasma carnitine ester profile in adult celiac disease patients maintained on long-term gluten free diet. *World J Gastroenterol,* 11**,** 6671-5.
BENUS, R. F., VAN DER WERF, T. S., WELLING, G. W., JUDD, P. A., TAYLOR, M. A., HARMSEN, H. J. & WHELAN, K. 2010. Association between Faecalibacterium prausnitzii and dietary fibre in colonic fermentation in healthy human subjects. *Br J Nutr,* 104**,** 693-700.
BEREZIN, A. E., KREMZER, A. A., MARTOVITSKAYA, Y. V., SAMURA, T. A., BEREZINA, T. A., ZULLI, A., KLIMAS, J. & KRUZLIAK, P. 2015. The utility of biomarker risk prediction score in patients with chronic heart failure. *Int J Clin Exp Med,* 8**,** 18255-64.
BERGE, J. M. 2009. A review of familial correlates of child and adolescent obesity: what has the 21st century taught us so far? *Int J Adolesc Med Health,* 21**,** 457-83.
BINGHAM, S. A., GILL, C., WELCH, A., CASSIDY, A., RUNSWICK, S. A., OAKES, S., LUBIN, R., THURNHAM, D. I., KEY, T. J., ROE, L., KHAW, K. T. & DAY, N. E. 1997. Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *Int J Epidemiol,* 26 Suppl 1**,** S137-51.
BINGHAM, S. A., WELCH, A. A., MCTAGGART, A., MULLIGAN, A. A., RUNSWICK, S. A., LUBEN, R., OAKES, S., KHAW, K. T., WAREHAM, N. & DAY, N. E. 2001. Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. *Public Health Nutr,* 4**,** 847-58.
BOGL, L. H., KAYE, S. M., RAMO, J. T., KANGAS, A. J., SOININEN, P., HAKKARAINEN, A., LUNDBOM, J., LUNDBOM, N., ORTEGA-ALONSO, A., RISSANEN, A., ALA-KORPELA, M., KAPRIO, J. & PIETILAINEN, K. H. 2016. Abdominal obesity and circulating metabolites: A twin study approach. *Metabolism,* 65**,** 111-21.
BONDIA-PONS, I., BARRI, T., HANHINEVA, K., JUNTUNEN, K., DRAGSTED, L. O., MYKKANEN, H. & POUTANEN, K. 2013. UPLC-QTOF/MS metabolic profiling unveils urinary changes in humans after a whole grain rye versus refined wheat bread intervention. *Mol Nutr Food Res,* 57**,** 412-22.
BONDIA-PONS, I., MAUKONEN, J., MATTILA, I., RISSANEN, A., SAARELA, M., KAPRIO, J., HAKKARAINEN, A., LUNDBOM, J., LUNDBOM, N., HYOTYLAINEN, T., PIETILAINEN, K. H. & ORESIC, M. 2014. Metabolome and fecal microbiota in monozygotic twin pairs discordant for weight: a Big Mac challenge. *Faseb j,* 28**,** 4169-79.
BONNEFOND, S., CATROUX, M., MELENOTTE, C., KARKOWSKI, L., ROLLAND, L., TROUILLIER, S. & RAFFRAY, L. 2016. Clinical features of actinomycosis: A retrospective, multicenter study of 28 cases of miscellaneous presentations. *Medicine (Baltimore),* 95**,** e3923.
BOUCHARD-MERCIER, A., PARADIS, A. M., PERUSSE, L. & VOHL, M. C. 2012. Associations between polymorphisms in genes involved in fatty acid metabolism and dietary fat intakes. *J Nutrigenet Nutrigenomics,* 5**,** 1-12.
BOURSIER, J., MUELLER, O., BARRET, M., MACHADO, M., FIZANNE, L., ARAUJO-PEREZ, F., GUY, C. D., SEED, P. C., RAWLS, J. F., DAVID, L. A., HUNAULT, G., OBERTI, F., CALES, P. & DIEHL, A. M. 2016. The severity of nonalcoholic fatty liver disease is associated with gut dysbiosis and shift in the metabolic function of the gut microbiota. *Hepatology,* 63**,** 764-75.
BREEN, F. M., PLOMIN, R. & WARDLE, J. 2006. Heritability of food preferences in young children. *Physiol Behav,* 88**,** 443-7.
BREMER, J. 1983. Carnitine--metabolism and functions. *Physiol Rev,* 63**,** 1420-80.
BRESLIN, P. A. 2013. An evolutionary perspective on food and human taste. *Curr Biol,* 23**,** R409-18.
BROWN, J. E. & CARLSON, M. 2000. Nutrition and multifetal pregnancy. *J Am Diet Assoc,* 100**,** 343-8.
BURMESTER, T., GERLACH, F. & HANKELN, T. 2007. Regulation and role of neuroglobin and cytoglobin under hypoxia. *Adv Exp Med Biol,* 618**,** 169-80.
BURMESTER, T., WEICH, B., REINHARDT, S. & HANKELN, T. 2000. A vertebrate globin expressed in the brain. *Nature,* 407**,** 520-3.
BUTLER, M. A., LANG, N. P., YOUNG, J. F., CAPORASO, N. E., VINEIS, P., HAYES, R. B., TEITEL, C. H., MASSENGILL, J. P., LAWSEN, M. F. & KADLUBAR, F. F. 1992.

Determination of CYP1A2 and NAT2 phenotypes in human populations by analysis of caffeine urinary metabolites. *Pharmacogenetics,* 2**,** 116-27.
BUTTE, N. F., LIU, Y., ZAKERI, I. F., MOHNEY, R. P., MEHTA, N., VORUGANTI, V. S., GORING, H., COLE, S. A. & COMUZZIE, A. G. 2015. Global metabolomic profiling targeting childhood obesity in the Hispanic population. *Am J Clin Nutr,* 102**,** 256-67.
CALVANI, R., MICCHELI, A., CAPUANI, G., TOMASSINI MICCHELI, A., PUCCETTI, C., DELFINI, M., IACONELLI, A., NANNI, G. & MINGRONE, G. 2010. Gut microbiome-derived metabolites characterize a peculiar obese urinary metabotype. *Int J Obes (Lond),* 34**,** 1095-8.
CAPORASO, J. G., KUCZYNSKI, J., STOMBAUGH, J., BITTINGER, K., BUSHMAN, F. D., COSTELLO, E. K., FIERER, N., PENA, A. G., GOODRICH, J. K., GORDON, J. I., HUTTLEY, G. A., KELLEY, S. T., KNIGHTS, D., KOENIG, J. E., LEY, R. E., LOZUPONE, C. A., MCDONALD, D., MUEGGE, B. D., PIRRUNG, M., REEDER, J., SEVINSKY, J. R., TURNBAUGH, P. J., WALTERS, W. A., WIDMANN, J., YATSUNENKO, T., ZANEVELD, J. & KNIGHT, R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods,* 7**,** 335-6.
CARON-JOBIN, M., MORISSET, A. S., TREMBLAY, A., HUOT, C., LEGARE, D. & TCHERNOF, A. 2011. Elevated serum 25(OH)D concentrations, vitamin D, and calcium intakes are associated with reduced adipocyte size in women. *Obesity (Silver Spring),* 19**,** 1335-41.
CHARLTON, K. E., TAPSELL, L. C., BATTERHAM, M. J., O'SHEA, J., THORNE, R., BECK, E. & TOSH, S. M. 2012. Effect of 6 weeks' consumption of beta-glucan-rich oat products on cholesterol levels in mildly hypercholesterolaemic overweight adults. *Br J Nutr,* 107**,** 1037-47.
CHRISTOPHERSON, M. R., DAWSON, J. A., STEVENSON, D. M., CUNNINGHAM, A. C., BRAMHACHARYA, S., WEIMER, P. J., KENDZIORSKI, C. & SUEN, G. 2014. Unique aspects of fiber degradation by the ruminal ethanologen Ruminococcus albus 7 revealed by physiological and transcriptomic analysis. *BMC Genomics,* 15**,** 1066.
CLAESSON, M. J., JEFFERY, I. B., CONDE, S., POWER, S. E., O'CONNOR, E. M., CUSACK, S., HARRIS, H. M., COAKLEY, M., LAKSHMINARAYANAN, B., O'SULLIVAN, O., FITZGERALD, G. F., DEANE, J., O'CONNOR, M., HARNEDY, N., O'CONNOR, K., O'MAHONY, D., VAN SINDEREN, D., WALLACE, M., BRENNAN, L., STANTON, C., MARCHESI, J. R., FITZGERALD, A. P., SHANAHAN, F., HILL, C., ROSS, R. P. & O'TOOLE, P. W. 2012. Gut microbiota composition correlates with diet and health in the elderly. *Nature,* 488**,** 178-84.
COLLINS, M. D. & GIBSON, G. R. 1999. Probiotics, prebiotics, and synbiotics: approaches for modulating the microbial ecology of the gut. *Am J Clin Nutr,* 69**,** 1052s-1057s.
COOKE, L. J., HAWORTH, C. M. & WARDLE, J. 2007. Genetic and environmental influences on children's food neophobia. *Am J Clin Nutr,* 86**,** 428-33.
COOPER, A. J., SHARP, S. J., LUBEN, R. N., KHAW, K. T., WAREHAM, N. J. & FOROUHI, N. G. 2015. The association between a biomarker score for fruit and vegetable intake and incident type 2 diabetes: the EPIC-Norfolk study. *Eur J Clin Nutr,* 69**,** 449-54.
COTILLARD, A., KENNEDY, S. P., KONG, L. C., PRIFTI, E., PONS, N., LE CHATELIER, E., ALMEIDA, M., QUINQUIS, B., LEVENEZ, F., GALLERON, N., GOUGIS, S., RIZKALLA, S., BATTO, J. M., RENAULT, P., DORE, J., ZUCKER, J. D., CLEMENT, K. & EHRLICH, S. D. 2013. Dietary intervention impact on gut microbial gene richness. *Nature,* 500**,** 585-8.
DAL MOLIN NETTO, B., LANDI MASQUIO, D. C., DA SILVEIRA CAMPOS, R. M., DE LIMA SANCHES, P., CAMPOS CORGOSINHO, F., TOCK, L., MISSAE OYAMA, L., TULIO DE MELLO, M., TUFIK, S. & DAMASO, A. R. 2014. The high glycemic index diet was an independent predictor to explain changes in agouti-related protein in obese adolescents. *Nutr Hosp,* 29**,** 305-14.
DASH, C., GOODMAN, M., FLANDERS, W. D., MINK, P. J., MCCULLOUGH, M. L. & BOSTICK, R. M. 2013. Using pathway-specific comprehensive exposure scores in epidemiology: application to oxidative balance in a pooled case-control study of incident, sporadic colorectal adenomas. *Am J Epidemiol,* 178**,** 610-24.
DAVID, L. A., MAURICE, C. F., CARMODY, R. N., GOOTENBERG, D. B., BUTTON, J. E., WOLFE, B. E., LING, A. V., DEVLIN, A. S., VARMA, Y., FISCHBACH, M. A., BIDDINGER, S. B., DUTTON, R. J. & TURNBAUGH, P. J. 2014. Diet rapidly and reproducibly alters the human gut microbiome. *Nature,* 505**,** 559-63.
DE CASTRO, J. M. 1993a. Genetic influences on daily intake and meal patterns of humans. *Physiol Behav,* 53**,** 777-82.

DE CASTRO, J. M. 1993b. Independence of genetic influences on body size, daily intake, and meal patterns of humans. *Physiol Behav,* 54**,** 633-9.
DE CASTRO, J. M. 2006. Heredity influences the dietary energy density of free-living humans. *Physiol Behav,* 87**,** 192-8.
DEHAVEN, C. D., EVANS, A. M., DAI, H. & LAWTON, K. A. 2010. Organization of GC/MS and LC/MS metabolomics data into chemical libraries. *J Cheminform,* 2**,** 9.
DEIANA, M., ROSA, A., CASU, V., PIGA, R., ASSUNTA DESSI, M. & ARUOMA, O. I. 2004. L-ergothioneine modulates oxidative damage in the kidney and liver of rats in vivo: studies upon the profile of polyunsaturated fatty acids. *Clin Nutr,* 23**,** 183-93.
DELANGHE, J., DE SLYPERE, J. P., DE BUYZERE, M., ROBBRECHT, J., WIEME, R. & VERMEULEN, A. 1989. Normal reference values for creatine, creatinine, and carnitine are lower in vegetarians. *Clin Chem,* 35**,** 1802-3.
DEMMER, R. T., JACOBS, D. R., JR., SINGH, R., ZUK, A., ROSENBAUM, M., PAPAPANOU, P. N. & DESVARIEUX, M. 2015. Periodontal Bacteria and Prediabetes Prevalence in ORIGINS: The Oral Infections, Glucose Intolerance, and Insulin Resistance Study. *J Dent Res,* 94**,** 201s-11s.
DESPRES, J. P., MOORJANI, S., LUPIEN, P. J., TREMBLAY, A., NADEAU, A. & BOUCHARD, C. 1992. Genetic aspects of susceptibility to obesity and related dyslipidemias. *Mol Cell Biochem,* 113**,** 151-69.
DIREK, K., CECELJA, M., ASTLE, W., CHOWIENCZYK, P., SPECTOR, T. D., FALCHI, M. & ANDREW, T. 2013. The relationship between DXA-based and anthropometric measures of visceral fat and morbidity in women. *BMC Cardiovasc Disord,* 13**,** 25.
DREHMER, M., PEREIRA, M. A., SCHMIDT, M. I., ALVIM, S., LOTUFO, P. A., LUFT, V. C. & DUNCAN, B. B. 2016. Total and Full-Fat, but Not Low-Fat, Dairy Product Intakes are Inversely Associated with Metabolic Syndrome in Adults. *J Nutr,* 146**,** 81-9.
DUBOIS, L., OHM KYVIK, K., GIRARD, M., TATONE-TOKUDA, F., PERUSSE, D., HJELMBORG, J., SKYTTHE, A., RASMUSSEN, F., WRIGHT, M. J., LICHTENSTEIN, P. & MARTIN, N. G. 2012. Genetic and environmental contributions to weight, height, and BMI from birth to 19 years of age: an international study of over 12,000 twin pairs. *PLoS One,* 7**,** e30153.
DUFFY, V. B., LANIER, S. A., HUTCHINS, H. L., PESCATELLO, L. S., JOHNSON, M. K. & BARTOSHUK, L. M. 2007. Food preference questionnaire as a screening tool for assessing dietary risk of cardiovascular disease within health risk appraisals. *J Am Diet Assoc,* 107**,** 237-45.
DUNCAN, S. H., LOBLEY, G. E., HOLTROP, G., INCE, J., JOHNSTONE, A. M., LOUIS, P. & FLINT, H. J. 2008. Human colonic microbiota associated with diet, obesity and weight loss. *Int J Obes (Lond),* 32**,** 1720-4.
EADY, J. J., WORTLEY, G. M., WORMSTONE, Y. M., HUGHES, J. C., ASTLEY, S. B., FOXALL, R. J., DOLEMAN, J. F. & ELLIOTT, R. M. 2005. Variation in gene expression profiles of peripheral blood mononuclear cells from healthy volunteers. *Physiol Genomics,* 22**,** 402-11.
EDGAR, R. C., HAAS, B. J., CLEMENTE, J. C., QUINCE, C. & KNIGHT, R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics,* 27**,** 2194-200.
ESCOBAR, J. S., KLOTZ, B., VALDES, B. E. & AGUDELO, G. M. 2014. The gut microbiota of Colombians differs from that of Americans, Europeans and Asians. *BMC Microbiol,* 14**,** 311.
ETXEBERRIA, U., ARIAS, N., BOQUE, N., MACARULLA, M. T., PORTILLO, M. P., MILAGRO, F. I. & MARTINEZ, J. A. 2015. Shifts in microbiota species and fermentation products in a dietary model enriched in fat and sucrose. *Benef Microbes,* 6**,** 97-111.
EY, J., SCHOMIG, E. & TAUBERT, D. 2007. Dietary sources and antioxidant effects of ergothioneine. *J Agric Food Chem,* 55**,** 6466-74.
FAITH, D. P. 1992. CONSERVATION EVALUATION AND PHYLOGENETIC DIVERSITY. *Biological Conservation,* 61**,** 1-10.
FAITH, M. S., RHA, S. S., NEALE, M. C. & ALLISON, D. B. 1999. Evidence for genetic influences on human energy intake: results from a twin study using measured observations. *Behav Genet,* 29**,** 145-54.
FAITH, M. S., RHEA, S. A., CORLEY, R. P. & HEWITT, J. K. 2008. Genetic and shared environmental influences on children's 24-h food and beverage intake: sex differences at age 7 y. *Am J Clin Nutr,* 87**,** 903-11.
FEENEY, E., O'BRIEN, S., SCANNELL, A., MARKEY, A. & GIBNEY, E. R. 2011. Genetic variation in taste perception: does it have a role in healthy eating? *Proc Nutr Soc,* 70**,** 135-43.

FENG, X., LIU, X., LUO, Q. & LIU, B. F. 2008. Mass spectrometry in systems biology: an overview. *Mass Spectrom Rev,* 27**,** 635-60.
FERNANDES, J., SU, W., RAHAT-ROZENBLOOM, S., WOLEVER, T. M. & COMELLI, E. M. 2014. Adiposity, gut microbiota and faecal short chain fatty acids are linked in adult humans. *Nutr Diabetes,* 4**,** e121.
FILDES, A., VAN JAARSVELD, C. H., LLEWELLYN, C. H., FISHER, A., COOKE, L. & WARDLE, J. 2014. Nature and nurture in children's food preferences. *Am J Clin Nutr*.
FISCHER, J. E., BACHMANN, L. M. & JAESCHKE, R. 2003. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med,* 29**,** 1043-51.
FISCHER, K., MOEWES, D., KOCH, M., MULLER, H. P., JACOBS, G., KASSUBEK, J., LIEB, W. & NOTHLINGS, U. 2015. MRI-determined total volumes of visceral and subcutaneous abdominal and trunk adipose tissue are differentially and sex-dependently associated with patterns of estimated usual nutrient intake in a northern German population. *Am J Clin Nutr,* 101**,** 794-807.
FLOEGEL, A., STEFAN, N., YU, Z., MUHLENBRUCH, K., DROGAN, D., JOOST, H. G., FRITSCHE, A., HARING, H. U., HRABE DE ANGELIS, M., PETERS, A., RODEN, M., PREHN, C., WANG-SATTLER, R., ILLIG, T., SCHULZE, M. B., ADAMSKI, J., BOEING, H. & PISCHON, T. 2013a. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes,* 62**,** 639-48.
FLOEGEL, A., VON RUESTEN, A., DROGAN, D., SCHULZE, M. B., PREHN, C., ADAMSKI, J., PISCHON, T. & BOEING, H. 2013b. Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam. *Eur J Clin Nutr,* 67**,** 1100-8.
FRANCESCHINI, N., HAACK, K., ALMASY, L., LASTON, S., LEE, E. T., BEST, L. G., FABSITZ, R. R., MACCLUER, J. W., HOWARD, B. V., UMANS, J. G. & COLE, S. A. 2014. Generalization of associations of kidney-related genetic loci to American Indians. *Clin J Am Soc Nephrol,* 9**,** 150-8.
FRANKENFIELD, D. C., MUTH, E. R. & ROWE, W. A. 1998. The Harris-Benedict studies of human basal metabolism: history and limitations. *J Am Diet Assoc,* 98**,** 439-45.
GALL, W. E., BEEBE, K., LAWTON, K. A., ADAM, K. P., MITCHELL, M. W., NAKHLE, P. J., RYALS, J. A., MILBURN, M. V., NANNIPIERI, M., CAMASTRA, S., NATALI, A. & FERRANNINI, E. 2010. alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. *PLoS One,* 5**,** e10883.
GATLEY, S. J. & SHERRATT, H. S. 1977. The synthesis of hippurate from benzoate and glycine by rat liver mitochondria. Submitochondrial localization and kinetics. *Biochem J,* 166**,** 39-47.
GIACCO, R., CLEMENTE, G., CIPRIANO, D., LUONGO, D., VISCOVO, D., PATTI, L., DI MARINO, L., GIACCO, A., NAVIGLIO, D., BIANCHI, M. A., CIATI, R., BRIGHENTI, F., RIVELLESE, A. A. & RICCARDI, G. 2010. Effects of the regular consumption of wholemeal wheat foods on cardiovascular risk factors in healthy people. *Nutr Metab Cardiovasc Dis,* 20**,** 186-94.
GIBBONS, H., MCNULTY, B. A., NUGENT, A. P., WALTON, J., FLYNN, A., GIBNEY, M. J. & BRENNAN, L. 2015. A metabolomics approach to the identification of biomarkers of sugar-sweetened beverage intake. *Am J Clin Nutr,* 101**,** 471-7.
GIBNEY, M. J., WALSH, M., BRENNAN, L., ROCHE, H. M., GERMAN, B. & VAN OMMEN, B. 2005. Metabolomics in human nutrition: opportunities and challenges. *Am J Clin Nutr,* 82**,** 497-503.
GLASS, C. K. & WITZTUM, J. L. 2001. Atherosclerosis. the road ahead. *Cell,* 104**,** 503-16.
GOLLEY, R. K., MCNAUGHTON, S. A. & HENDRIE, G. A. 2015. A dietary guideline adherence score is positively associated with dietary biomarkers but not lipid profile in healthy children. *J Nutr,* 145**,** 128-33.
GONCALVES, I., EDSFELDT, A., KO, N. Y., GRUFMAN, H., BERG, K., BJORKBACKA, H., NITULESCU, M., PERSSON, A., NILSSON, M., PREHN, C., ADAMSKI, J. & NILSSON, J. 2012. Evidence supporting a key role of Lp-PLA2-generated lysophosphatidylcholine in human atherosclerotic plaque inflammation. *Arterioscler Thromb Vasc Biol,* 32**,** 1505-12.
GONTHIER, M. P., VERNY, M. A., BESSON, C., REMESY, C. & SCALBERT, A. 2003. Chlorogenic acid bioavailability largely depends on its metabolism by the gut microflora in rats. *J Nutr,* 133**,** 1853-9.
GOODRICH, J. K., DAVENPORT, E. R., BEAUMONT, M., JACKSON, M. A., KNIGHT, R., OBER, C., SPECTOR, T. D., BELL, J. T., CLARK, A. G. & LEY, R. E. 2016. Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host Microbe,* 19**,** 731-43.

GOODRICH, J. K., WATERS, J. L., POOLE, A. C., SUTTER, J. L., KOREN, O., BLEKHMAN, R., BEAUMONT, M., VAN TREUREN, W., KNIGHT, R., BELL, J. T., SPECTOR, T. D., CLARK, A. G. & LEY, R. E. 2014a. Human genetics shape the gut microbiome. *Cell,* 159**,** 789-99.

GOODRICH, JULIA K., WATERS, JILLIAN L., POOLE, ANGELA C., SUTTER, JESSICA L., KOREN, O., BLEKHMAN, R., BEAUMONT, M., VAN TREUREN, W., KNIGHT, R., BELL, JORDANA T., SPECTOR, TIMOTHY D., CLARK, ANDREW G. & LEY, RUTH E. 2014b. Human Genetics Shape the Gut Microbiome. *Cell,* 159**,** 789-799.

GRUNDBERG, E., SMALL, K. S., HEDMAN, A. K., NICA, A. C., BUIL, A., KEILDSON, S., BELL, J. T., YANG, T. P., MEDURI, E., BARRETT, A., NISBETT, J., SEKOWSKA, M., WILK, A., SHIN, S. Y., GLASS, D., TRAVERS, M., MIN, J. L., RING, S., HO, K., THORLEIFSSON, G., KONG, A., THORSTEINDOTTIR, U., AINALI, C., DIMAS, A. S., HASSANALI, N., INGLE, C., KNOWLES, D., KRESTYANINOVA, M., LOWE, C. E., DI MEGLIO, P., MONTGOMERY, S. B., PARTS, L., POTTER, S., SURDULESCU, G., TSAPROUNI, L., TSOKA, S., BATAILLE, V., DURBIN, R., NESTLE, F. O., O'RAHILLY, S., SORANZO, N., LINDGREN, C. M., ZONDERVAN, K. T., AHMADI, K. R., SCHADT, E. E., STEFANSSON, K., SMITH, G. D., MCCARTHY, M. I., DELOUKAS, P., DERMITZAKIS, E. T. & SPECTOR, T. D. 2012. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet,* 44**,** 1084-9.

GUENTHER, P. M., CASAVALE, K. O., REEDY, J., KIRKPATRICK, S. I., HIZA, H. A., KUCZYNSKI, K. J., KAHLE, L. L. & KREBS-SMITH, S. M. 2013. Update of the Healthy Eating Index: HEI-2010. *J Acad Nutr Diet,* 113**,** 569-80.

GUERTIN, K. A., LOFTFIELD, E., BOCA, S. M., SAMPSON, J. N., MOORE, S. C., XIAO, Q., HUANG, W. Y., XIONG, X., FREEDMAN, N. D., CROSS, A. J. & SINHA, R. 2015. Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer. *Am J Clin Nutr,* 101**,** 1000-11.

GUERTIN, K. A., MOORE, S. C., SAMPSON, J. N., HUANG, W. Y., XIAO, Q., STOLZENBERG-SOLOMON, R. Z., SINHA, R. & CROSS, A. J. 2014a. Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. *Am J Clin Nutr,* 100**,** 208-17.

GUERTIN, K. A., MOORE, S. C., SAMPSON, J. N., HUANG, W. Y., XIAO, Q., STOLZENBERG-SOLOMON, R. Z., SINHA, R. & CROSS, A. J. 2014b. Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. *Am J Clin Nutr,* 100**,** 208-217.

GUNDERSON, E. P., TSAI, A. L., SELBY, J. V., CAAN, B., MAYER-DAVIS, E. J. & RISCH, N. 2006. Twins of mistaken zygosity (TOMZ): evidence for genetic contributions to dietary patterns and physiologic traits. *Twin Res Hum Genet,* 9**,** 540-9.

GUYENET, S. J. & SCHWARTZ, M. W. 2012. Clinical review: Regulation of food intake, energy balance, and body fat mass: implications for the pathogenesis and treatment of obesity. *J Clin Endocrinol Metab,* 97**,** 745-55.

HAIRSTON, K. G., VITOLINS, M. Z., NORRIS, J. M., ANDERSON, A. M., HANLEY, A. J. & WAGENKNECHT, L. E. 2012. Lifestyle factors and 5-year abdominal fat accumulation in a minority cohort: the IRAS Family Study. *Obesity (Silver Spring),* 20**,** 421-7.

HARO, C., GARCIA-CARPINTERO, S., ALCALA-DIAZ, J. F., GOMEZ-DELGADO, F., DELGADO-LISTA, J., PEREZ-MARTINEZ, P., RANGEL ZUNIGA, O. A., QUINTANA-NAVARRO, G. M., LANDA, B. B., CLEMENTE, J. C., LOPEZ-MIRANDA, J., CAMARGO, A. & PEREZ-JIMENEZ, F. 2016. The gut microbial community in metabolic syndrome patients is modified by diet. *J Nutr Biochem,* 27**,** 27-31.

HASSELBALCH, A. L., ANGQUIST, L., CHRISTIANSEN, L., HEITMANN, B. L., KYVIK, K. O. & SORENSEN, T. I. 2010. A variant in the fat mass and obesity-associated gene (FTO) and variants near the melanocortin-4 receptor gene (MC4R) do not influence dietary intake. *J Nutr,* 140**,** 831-4.

HASSELBALCH, A. L., HEITMANN, B. L., KYVIK, K. O. & SORENSEN, T. I. 2008. Studies of twins indicate that genetics influence dietary intake. *J Nutr,* 138**,** 2406-12.

HEALTH., D. O. 1991. *Dietary reference values for food energy and nutrients for the United Kingdom. Report of the panel on dietary reference values of the Committee on Medical Aspects of Food Policy. Report on Health and Social Subjects 41.,* London, HMSO.

HELLER, R. F., O'CONNELL, D. L., ROBERTS, D. C., ALLEN, J. R., KNAPP, J. C., STEELE, P. L. & SILOVE, D. 1988. Lifestyle factors in monozygotic and dizygotic twins. *Genet Epidemiol,* 5**,** 311-21.

HOFFMAN, R. & GERBER, M. 2013. Evaluating and adapting the Mediterranean diet for non-Mediterranean populations: a critical appraisal. *Nutr Rev,* 71**,** 573-84.
HOLLAENDER, P. L., ROSS, A. B. & KRISTENSEN, M. 2015. Whole-grain and blood lipid changes in apparently healthy adults: a systematic review and meta-analysis of randomized controlled studies. *Am J Clin Nutr,* 102**,** 556-72.
HOLLE, R., HAPPICH, M., LOWEL, H. & WICHMANN, H. E. 2005. KORA--a research platform for population based health research. *Gesundheitswesen,* 67 Suppl 1**,** S19-25.
HSIAO, E. Y., MCBRIDE, S. W., HSIEN, S., SHARON, G., HYDE, E. R., MCCUE, T., CODELLI, J. A., CHOW, J., REISMAN, S. E., PETROSINO, J. F., PATTERSON, P. H. & MAZMANIAN, S. K. 2013. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell,* 155**,** 1451-63.
HU, F. B. 2002. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol,* 13**,** 3-9.
HUANG, X., SJOGREN, P., ARNLOV, J., CEDERHOLM, T., LIND, L., STENVINKEL, P., LINDHOLM, B., RISERUS, U. & CARRERO, J. J. 2014. Serum fatty acid patterns, insulin sensitivity and the metabolic syndrome in individuals with chronic kidney disease. *J Intern Med,* 275**,** 71-83.
HUGHES, M. F., SAARELA, O., BLANKENBERG, S., ZELLER, T., HAVULINNA, A. S., KUULASMAA, K., YARNELL, J., SCHNABEL, R. B., TIRET, L., SALOMAA, V., EVANS, A. & KEE, F. 2012. A multiple biomarker risk score for guiding clinical decisions using a decision curve approach. *Eur J Prev Cardiol,* 19**,** 874-84.
HUR, Y. M., BOUCHARD, T. J., JR. & ECKERT, E. 1998. Genetic and environmental influences on self-reported diet: a reared-apart twin study. *Physiol Behav,* 64**,** 629-36.
ILLIG, T., GIEGER, C., ZHAI, G., ROMISCH-MARGL, W., WANG-SATTLER, R., PREHN, C., ALTMAIER, E., KASTENMULLER, G., KATO, B. S., MEWES, H. W., MEITINGER, T., DE ANGELIS, M. H., KRONENBERG, F., SORANZO, N., WICHMANN, H. E., SPECTOR, T. D., ADAMSKI, J. & SUHRE, K. 2010. A genome-wide perspective of genetic variation in human metabolism. *Nat Genet,* 42**,** 137-41.
IMHANN, F., BONDER, M. J., VICH VILA, A., FU, J., MUJAGIC, Z., VORK, L., TIGCHELAAR, E. F., JANKIPERSADSING, S. A., CENIT, M. C., HARMSEN, H. J., DIJKSTRA, G., FRANKE, L., XAVIER, R. J., JONKERS, D., WIJMENGA, C., WEERSMA, R. K. & ZHERNAKOVA, A. 2016. Proton pump inhibitors affect the gut microbiome. *Gut,* 65**,** 740-8.
JACKSON, M. A., BELL, J. T., SPECTOR, T. & STEVES, C. 2016a. A heritability-based comparison of methods used to cluster 16S rRNA gene sequences into operational taxonomic units. *Peer J Preprints*.
JACKSON, M. A., JEFFERY, I. B., BEAUMONT, M., BELL, J. T., CLARK, A. G., LEY, R. E., O'TOOLE, P. W., SPECTOR, T. D. & STEVES, C. J. 2016b. Signatures of early frailty in the gut microbiota. *Genome Med,* 8**,** 8.
JEFFERY, I. B., CLAESSON, M. J., O'TOOLE, P. W. & SHANAHAN, F. 2012. Categorization of the gut microbiota: enterotypes or gradients? *Nat Rev Microbiol,* 10**,** 591-2.
JENKO-PRAZNIKAR, Z., PETELIN, A., JURDANA, M. & ZIBERNA, L. 2013. Serum bilirubin levels are lower in overweight asymptomatic middle-aged adults: an early indicator of metabolic syndrome? *Metabolism,* 62**,** 976-85.
JENSEN, B. M., ADHIKARI, K. B., SCHNOOR, H. J., JUEL-BERG, N., FOMSGAARD, I. S. & POULSEN, L. K. 2015. Quantitative analysis of absorption, metabolism, and excretion of benzoxazinoids in humans after the consumption of high- and low-benzoxazinoid diets with similar contents of cereal dietary fibres: a crossover study. *Eur J Nutr*.
JOHNSON, I. T., GEE, J. M., PRICE, K., CURL, C. & FENWICK, G. R. 1986. Influence of saponins on gut permeability and active nutrient transport in vitro. *J Nutr,* 116**,** 2270-7.
KALHAN, S. C., GUO, L., EDMISON, J., DASARATHY, S., MCCULLOUGH, A. J., HANSON, R. W. & MILBURN, M. 2011. Plasma metabolomic profile in nonalcoholic fatty liver disease. *Metabolism,* 60**,** 404-13.
KARBOWNIK, M., REITER, R. J., GARCIA, J. J., CABRERA, J., BURKHARDT, S., OSUNA, C. & LEWINSKI, A. 2001. Indole-3-propionic acid, a melatonin-related molecule, protects hepatic microsomal membranes from iron-induced oxidative damage: relevance to cancer reduction. *J Cell Biochem,* 81**,** 507-13.
KASTORINI, C. M., MILIONIS, H. J., ESPOSITO, K., GIUGLIANO, D., GOUDEVENOS, J. A. & PANAGIOTAKOS, D. B. 2011. The effect of Mediterranean diet on metabolic syndrome and its components: a meta-analysis of 50 studies and 534,906 individuals. *J Am Coll Cardiol,* 57**,** 1299-313.

KASUYA, F., IGARASHI, K. & FUKUI, M. 1996. Participation of a medium chain acyl-CoA synthetase in glycine conjugation of the benzoic acid derivatives with the electron-donating groups. *Biochem Pharmacol,* 51**,** 805-9.

KESKITALO, K., SILVENTOINEN, K., TUORILA, H., PEROLA, M., PIETILAINEN, K. H., RISSANEN, A. & KAPRIO, J. 2008. Genetic and environmental contributions to food use patterns of young adult twins. *Physiol Behav,* 93**,** 235-42.

KETTUNEN, J., TUKIAINEN, T., SARIN, A. P., ORTEGA-ALONSO, A., TIKKANEN, E., LYYTIKAINEN, L. P., KANGAS, A. J., SOININEN, P., WURTZ, P., SILANDER, K., DICK, D. M., ROSE, R. J., SAVOLAINEN, M. J., VIIKARI, J., KAHONEN, M., LEHTIMAKI, T., PIETILAINEN, K. H., INOUYE, M., MCCARTHY, M. I., JULA, A., ERIKSSON, J., RAITAKARI, O. T., SALOMAA, V., KAPRIO, J., JARVELIN, M. R., PELTONEN, L., PEROLA, M., FREIMER, N. B., ALA-KORPELA, M., PALOTIE, A. & RIPATTI, S. 2012. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet,* 44**,** 269-76.

KHO, M., LEE, J. E., SONG, Y. M., LEE, K., KIM, K., YANG, S., JOUNG, H. & SUNG, J. 2013. Genetic and environmental influences on sodium intake determined by using half-day urine samples: the Healthy Twin Study. *Am J Clin Nutr,* 98**,** 1410-6.

KIM, J. H., KIM, S. H., SONG, S. Y., KIM, W. S., SONG, S. U., YI, T., JEON, M. S., CHUNG, H. M., XIA, Y. & SUNG, J. H. 2014. Hypoxia induces adipocyte differentiation of adipose-derived stem cells by triggering reactive oxygen species generation. *Cell Biol Int,* 38**,** 32-40.

KIM, Y. S., XUN, P., IRIBARREN, C., VAN HORN, L., STEFFEN, L., DAVIGLUS, M. L., SISCOVICK, D., LIU, K. & HE, K. 2016. Intake of fish and long-chain omega-3 polyunsaturated fatty acids and incidence of metabolic syndrome among American young adults: a 25-year follow-up study. *Eur J Nutr,* 55**,** 1707-16.

KIYOHARA, C., WASHIO, M., HORIUCHI, T., ASAMI, T., IDE, S., ATSUMI, T., KOBASHI, G., TAKAHASHI, H. & TADA, Y. 2014. Modifying effect of N-acetyltransferase 2 genotype on the association between systemic lupus erythematosus and consumption of alcohol and caffeine-rich beverages. *Arthritis Care Res (Hoboken),* 66**,** 1048-56.

KNAAPILA, A., HWANG, L. D., LYSENKO, A., DUKE, F. F., FESI, B., KHOSHNEVISAN, A., JAMES, R. S., WYSOCKI, C. J., RHYU, M., TORDOFF, M. G., BACHMANOV, A. A., MURA, E., NAGAI, H. & REED, D. R. 2012. Genetic analysis of chemosensory traits in human twins. *Chem Senses,* 37**,** 869-81.

KNUDSEN, D., JUTFELT, F., SUNDH, H., SUNDELL, K., KOPPE, W. & FROKIAER, H. 2008. Dietary soya saponins increase gut permeability and play a key role in the onset of soyabean-induced enteritis in Atlantic salmon (Salmo salar L.). *Br J Nutr,* 100**,** 120-9.

KOETH, R. A., WANG, Z., LEVISON, B. S., BUFFA, J. A., ORG, E., SHEEHY, B. T., BRITT, E. B., FU, X., WU, Y., LI, L., SMITH, J. D., DIDONATO, J. A., CHEN, J., LI, H., WU, G. D., LEWIS, J. D., WARRIER, M., BROWN, J. M., KRAUSS, R. M., TANG, W. H., BUSHMAN, F. D., LUSIS, A. J. & HAZEN, S. L. 2013. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med,* 19**,** 576-85.

KOREN, O., SPOR, A., FELIN, J., FAK, F., STOMBAUGH, J., TREMAROLI, V., BEHRE, C. J., KNIGHT, R., FAGERBERG, B., LEY, R. E. & BACKHED, F. 2011. Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proc Natl Acad Sci U S A,* 108 Suppl 1**,** 4592-8.

KOSTIC, A. D., XAVIER, R. J. & GEVERS, D. 2014. The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology,* 146**,** 1489-99.

KRISHNAN, N., DICKMAN, M. B. & BECKER, D. F. 2008. Proline modulates the intracellular redox environment and protects mammalian cells against oxidative stress. *Free Radic Biol Med,* 44**,** 671-81.

KRUMSIEK, J., MITTELSTRASS, K., DO, K. T., STUCKLER, F., RIED, J., ADAMSKI, J., PETERS, A., ILLIG, T., KRONENBERG, F., FRIEDRICH, N., NAUCK, M., PIETZNER, M., MOOK-KANAMORI, D. O., SUHRE, K., GIEGER, C., GRALLERT, H., THEIS, F. J. & KASTENMULLER, G. 2015. Gender-specific pathway differences in the human serum metabolome. *Metabolomics,* 11**,** 1815-1833.

KUSSMANN, M., RAYMOND, F. & AFFOLTER, M. 2006. OMICS-driven biomarker discovery in nutrition and health. *J Biotechnol,* 124**,** 758-87.

KWON, K. M., KAM, J. H., KIM, M. Y., KIM, M. Y., CHUNG, C. H., KIM, J. K., LINTON, J. A., EOM, A., KOH, S. B. & KANG, H. T. 2011. Inverse association between total bilirubin and metabolic syndrome in rural korean women. *J Womens Health (Larchmt),* 20**,** 963-9.

KYVIK, K. 2000. Generalisability and assumptions of twin studies. *In:* SPECTOR, T. D., SNIEDER, H. & MACGREGOR, A. J. (eds.) *Advances in twin and sib-pair analysis. .* London: Greenwich Medical Media.

LAHTI, L., SALONEN, A., KEKKONEN, R. A., SALOJARVI, J., JALANKA-TUOVINEN, J., PALVA, A., ORESIC, M. & DE VOS, W. M. 2013. Associations between the human intestinal microbiota, Lactobacillus rhamnosus GG and serum lipids indicated by integrated analysis of high-throughput profiling data. *PeerJ,* 1**,** e32.

LE CHATELIER, E., NIELSEN, T., QIN, J., PRIFTI, E., HILDEBRAND, F., FALONY, G., ALMEIDA, M., ARUMUGAM, M., BATTO, J. M., KENNEDY, S., LEONARD, P., LI, J., BURGDORF, K., GRARUP, N., JORGENSEN, T., BRANDSLUND, I., NIELSEN, H. B., JUNCKER, A. S., BERTALAN, M., LEVENEZ, F., PONS, N., RASMUSSEN, S., SUNAGAWA, S., TAP, J., TIMS, S., ZOETENDAL, E. G., BRUNAK, S., CLEMENT, K., DORE, J., KLEEREBEZEM, M., KRISTIANSEN, K., RENAULT, P., SICHERITZ-PONTEN, T., DE VOS, W. M., ZUCKER, J. D., RAES, J., HANSEN, T., BORK, P., WANG, J., EHRLICH, S. D. & PEDERSEN, O. 2013. Richness of human gut microbiome correlates with metabolic markers. *Nature,* 500**,** 541-6.

LEE, Y. H. & SONG, G. G. 2012. Pathway analysis of genome-wide association studies on uric acid concentrations. *Hum Immunol,* 73**,** 805-10.

LEY, R. E., TURNBAUGH, P. J., KLEIN, S. & GORDON, J. I. 2006. Microbial ecology: human gut microbes associated with obesity. *Nature,* 444**,** 1022-3.

LI, R. W., YANG, C., SIT, A. S., KWAN, Y. W., LEE, S. M., HOI, M. P., CHAN, S. W., HAUSMAN, M., VANHOUTTE, P. M. & LEUNG, G. P. 2014. Uptake and protective effects of ergothioneine in human endothelial cells. *J Pharmacol Exp Ther,* 350**,** 691-700.

LI, Z., XU, J., JIANG, T., GE, Y., LIU, P., ZHANG, M., SU, Z., GAO, C., MA, C. & XU, P. 2016. Overexpression of transport proteins improves the production of 5-aminovalerate from l-lysine in Escherichia coli. *Sci Rep,* 6**,** 30884.

LINDNER, I., RUBIN, D., HELWIG, U., NITZ, I., HAMPE, J., SCHREIBER, S., SCHREZENMEIR, J. & DORING, F. 2006. The L513S polymorphism in medium-chain acyl-CoA synthetase 2 (MACS2) is associated with risk factors of the metabolic syndrome in a Caucasian study population. *Mol Nutr Food Res,* 50**,** 270-4.

LINDTNER, C., SCHERER, T., ZIELINSKI, E., FILATOVA, N., FASSHAUER, M., TONKS, N. K., PUCHOWICZ, M. & BUETTNER, C. 2013. Binge drinking induces whole-body insulin resistance by impairing hypothalamic insulin action. *Sci Transl Med,* 5**,** 170ra14.

LING, Y., VAN HERPT, T. T., VAN HOEK, M., DEHGHAN, A., HOFMAN, A., UITTERLINDEN, A. G., JIANG, S., LIEVERSE, A. G., BRAVENBOER, B., LU, D., VAN DUIJN, C. M., GAO, X. & SIJBRANDS, E. J. 2014. A genetic variant in SLC6A20 is associated with Type 2 diabetes in white-European and Chinese populations. *Diabet Med,* 31**,** 1350-6.

LIU, C. T., GARNAAS, M. K., TIN, A., KOTTGEN, A., FRANCESCHINI, N., PERALTA, C. A., DE BOER, I. H., LU, X., ATKINSON, E., DING, J., NALLS, M., SHRINER, D., CORESH, J., KUTLAR, A., BIBBINS-DOMINGO, K., SISCOVICK, D., AKYLBEKOVA, E., WYATT, S., ASTOR, B., MYCHALECKJY, J., LI, M., REILLY, M. P., TOWNSEND, R. R., ADEYEMO, A., ZONDERMAN, A. B., DE ANDRADE, M., TURNER, S. T., MOSLEY, T. H., HARRIS, T. B., ROTIMI, C. N., LIU, Y., KARDIA, S. L., EVANS, M. K., SHLIPAK, M. G., KRAMER, H., FLESSNER, M. F., DREISBACH, A. W., GOESSLING, W., CUPPLES, L. A., KAO, W. L. & FOX, C. S. 2011. Genetic association for renal traits among participants of African ancestry reveals new loci for renal function. *PLoS Genet,* 7**,** e1002264.

LIU, J., TUVBLAD, C., RAINE, A. & BAKER, L. 2013. Genetic and environmental influences on nutrient intake. *Genes Nutr,* 8**,** 241-52.

LIU, S., BENNETT, D. C., TUN, H. M., KIM, J. E., CHENG, K. M., ZHANG, H. & LEUNG, F. C. 2015. The effect of diet and host genotype on ceca microbiota of Japanese quail fed a cholesterol enriched diet. *Front Microbiol,* 6**,** 1092.

LLEWELLYN, C. H., TRZASKOWSKI, M., PLOMIN, R. & WARDLE, J. 2014. From modeling to measurement: Developmental trends in genetic influence on adiposity in childhood. *Obesity (Silver Spring)*.

LLOYD, A. J., BECKMANN, M., HALDAR, S., SEAL, C., BRANDT, K. & DRAPER, J. 2013. Data-driven strategy for the discovery of potential urinary biomarkers of habitual dietary exposure. *Am J Clin Nutr,* 97**,** 377-89.

LUSTGARTEN, M. S., PRICE, L. L., CHALE, A., PHILLIPS, E. M. & FIELDING, R. A. 2014. Branched chain amino acids are associated with muscle mass in functionally limited older adults. *J Gerontol A Biol Sci Med Sci,* 69**,** 717-24.

MA, J., SLOAN, M., FOX, C. S., HOFFMANN, U., SMITH, C. E., SALTZMAN, E., ROGERS, G. T., JACQUES, P. F. & MCKEOWN, N. M. 2014. Sugar-sweetened beverage consumption is associated with abdominal fat partitioning in healthy adults. *J Nutr,* 144**,** 1283-90.

MACLELLAN, W. R., WANG, Y. & LUSIS, A. J. 2012. Systems-based approaches to cardiovascular disease. *Nat Rev Cardiol,* 9**,** 172-84.

MAGNUSDOTTIR, O. K., LANDBERG, R., GUNNARSDOTTIR, I., CLOETENS, L., AKESSON, B., ROSQVIST, F., SCHWAB, U., HERZIG, K. H., HUKKANEN, J., SAVOLAINEN, M. J., BRADER, L., HERMANSEN, K., KOLEHMAINEN, M., POUTANEN, K., UUSITUPA, M., RISERUS, U. & THORSDOTTIR, I. 2014. Whole grain rye intake, reflected by a biomarker, is associated with favorable blood lipid outcomes in subjects with the metabolic syndrome--a randomized study. *PLoS One,* 9**,** e110827.

MAITI, S. & CHEN, G. 2015. Ethanol up-regulates phenol sulfotransferase (SULT1A1) and hydroxysteroid sulfotransferase (SULT2A1) in rat liver and intestine. *Arch Physiol Biochem,* 121**,** 68-74.

MANICHANH, C., RIGOTTIER-GOIS, L., BONNAUD, E., GLOUX, K., PELLETIER, E., FRANGEUL, L., NALIN, R., JARRIN, C., CHARDON, P., MARTEAU, P., ROCA, J. & DORE, J. 2006. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut,* 55**,** 205-11.

MANINI, P., DE PALMA, G., ANDREOLI, R., GOLDONI, M., POLI, D., LASAGNI, G. & MUTTI, A. 2003. [Urinary excretion of 4-vinyl phenol after experimental and occupational exposure to styrene]. *G Ital Med Lav Ergon,* 25 Suppl**,** 61-2.

MARION, D. 2013. An introduction to biological NMR spectroscopy. *Mol Cell Proteomics,* 12**,** 3006-25.

MARKLEY, J. L., BRUSCHWEILER, R., EDISON, A. S., EGHBALNIA, H. R., POWERS, R., RAFTERY, D. & WISHART, D. S. 2016. The future of NMR-based metabolomics. *Curr Opin Biotechnol,* 43**,** 34-40.

MARKLUND, M., MAGNUSDOTTIR, O. K., ROSQVIST, F., CLOETENS, L., LANDBERG, R., KOLEHMAINEN, M., BRADER, L., HERMANSEN, K., POUTANEN, K. S., HERZIG, K. H., HUKKANEN, J., SAVOLAINEN, M. J., DRAGSTED, L. O., SCHWAB, U., PAANANEN, J., UUSITUPA, M., AKESSON, B., THORSDOTTIR, I. & RISERUS, U. 2014. A dietary biomarker approach captures compliance and cardiometabolic effects of a healthy Nordic diet in individuals with metabolic syndrome. *J Nutr,* 144**,** 1642-9.

MARQUES, T. M., WALL, R., O'SULLIVAN, O., FITZGERALD, G. F., SHANAHAN, F., QUIGLEY, E. M., COTTER, P. D., CRYAN, J. F., DINAN, T. G., ROSS, R. P. & STANTON, C. 2015. Dietary trans-10, cis-12-conjugated linoleic acid alters fatty acid metabolism and microbiota composition in mice. *Br J Nutr,* 113**,** 728-38.

MARTIN, F. P., REZZI, S., PERE-TREPAT, E., KAMLAGE, B., COLLINO, S., LEIBOLD, E., KASTLER, J., REIN, D., FAY, L. B. & KOCHHAR, S. 2009. Metabolic effects of dark chocolate consumption on energy, gut microbiota, and stress-related metabolism in free-living subjects. *J Proteome Res,* 8**,** 5568-79.

MARTINEZ, I., LATTIMER, J. M., HUBACH, K. L., CASE, J. A., YANG, J., WEBER, C. G., LOUK, J. A., ROSE, D. J., KYUREGHIAN, G., PETERSON, D. A., HAUB, M. D. & WALTER, J. 2013. Gut microbiome composition is linked to whole grain-induced immunological improvements. *Isme j,* 7**,** 269-80.

MATIJASIC, B. B., OBERMAJER, T., LIPOGLAVSEK, L., GRABNAR, I., AVGUSTIN, G. & ROGELJ, I. 2013. Association of dietary type with fecal microbiota in vegetarians and omnivores in Slovenia. *Eur J Nutr*.

MCCANCE, R. A., WIDDOWSON, E. M., HOLLAND, B., WELCH, A., BUSS, D. H. & F., G. B. M. O. A. 1991. *McCance and Widdowson's the composition of foods*.

MCCRORY, M. A., FUSS, P. J., MCCALLUM, J. E., YAO, M., VINKEN, A. G., HAYS, N. P. & ROBERTS, S. B. 1999. Dietary variety within food groups: association with energy intake and body fatness in men and women. *Am J Clin Nutr,* 69**,** 440-7.

MCGUE, M. & BOUCHARD, T. J., JR. 1984. Adjustment of twin data for the effects of age and sex. *Behav Genet,* 14**,** 325-43.

MENNI, C., FAUMAN, E., ERTE, I., PERRY, J. R., KASTENMULLER, G., SHIN, S. Y., PETERSEN, A. K., HYDE, C., PSATHA, M., WARD, K. J., YUAN, W., MILBURN, M., PALMER, C. N., FRAYLING, T. M., TRIMMER, J., BELL, J. T., GIEGER, C., MOHNEY, R. P., BROSNAN, M. J., SUHRE, K., SORANZO, N. & SPECTOR, T. D. 2013a. Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. *Diabetes,* 62**,** 4270-6.

MENNI, C., KASTENMULLER, G., PETERSEN, A. K., BELL, J. T., PSATHA, M., TSAI, P. C., GIEGER, C., SCHULZ, H., ERTE, I., JOHN, S., BROSNAN, M. J., WILSON, S. G.,

TSAPROUNI, L., LIM, E. M., STUCKEY, B., DELOUKAS, P., MOHNEY, R., SUHRE, K., SPECTOR, T. D. & VALDES, A. M. 2013b. Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int J Epidemiol,* 42**,** 1111-9.

MENNI, C., MIGAUD, M., GLASTONBURY, C. A., BEAUMONT, M., NIKOLAOU, A., SMALL, K. S., BROSNAN, M. J., MOHNEY, R. P., SPECTOR, T. D. & VALDES, A. M. 2016. Metabolomic profiling to dissect the role of visceral fat in cardiometabolic health. *Obesity (Silver Spring),* 24**,** 1380-8.

MENNI, C., ZHAI, G., MACGREGOR, A., PREHN, C., ROMISCH-MARGL, W., SUHRE, K., ADAMSKI, J., CASSIDY, A., ILLIG, T., SPECTOR, T. D. & VALDES, A. M. 2013c. Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. *Metabolomics,* 9**,** 506-514.

METRUSTRY, S. J., EDWARDS, M. H., MEDLAND, S. E., HOLLOWAY, J. W., MONTGOMERY, G. W., MARTIN, N. G., SPECTOR, T. D., COOPER, C. & VALDES, A. M. 2014. Variants Close to NTRK2 Gene Are Associated With Birth Weight in Female Twins. *Twin Res Hum Genet***,** 1-8.

MICHAIL, S., DURBIN, M., TURNER, D., GRIFFITHS, A. M., MACK, D. R., HYAMS, J., LELEIKO, N., KENCHE, H., STOLFI, A. & WINE, E. 2012. Alterations in the gut microbiome of children with severe ulcerative colitis. *Inflamm Bowel Dis,* 18**,** 1799-808.

MILLS, C. E., TZOUNIS, X., ORUNA-CONCHA, M. J., MOTTRAM, D. S., GIBSON, G. R. & SPENCER, J. P. 2015. In vitro colonic metabolism of coffee and chlorogenic acid results in selective changes in human faecal microbiota growth. *Br J Nutr,* 113**,** 1220-7.

MOAYYERI, A., HAMMOND, C. J., HART, D. J. & SPECTOR, T. D. 2013. The UK Adult Twin Registry (TwinsUK Resource). *Twin Res Hum Genet,* 16**,** 144-9.

MOLLARD, R. C., SENECHAL, M., MACINTOSH, A. C., HAY, J., WICKLOW, B. A., WITTMEIER, K. D., SELLERS, E. A., DEAN, H. J., RYNER, L., BERARD, L. & MCGAVOCK, J. M. 2014. Dietary determinants of hepatic steatosis and visceral adiposity in overweight and obese youth at risk of type 2 diabetes. *Am J Clin Nutr,* 99**,** 804-12.

MOORE, S. C., MATTHEWS, C. E., SAMPSON, J. N., STOLZENBERG-SOLOMON, R. Z., ZHENG, W., CAI, Q., TAN, Y. T., CHOW, W. H., JI, B. T., LIU, D. K., XIAO, Q., BOCA, S. M., LEITZMANN, M. F., YANG, G., XIANG, Y. B., SINHA, R., SHU, X. O. & CROSS, A. J. 2014. Human metabolic correlates of body mass index. *Metabolomics,* 10**,** 259-269.

NAKAYAMA, A., MATSUO, H., SHIMIZU, T., OGATA, H., TAKADA, Y., NAKASHIMA, H., NAKAMURA, T., SHIMIZU, S., CHIBA, T., SAKIYAMA, M., USHIYAMA, C., TAKADA, T., INOUE, K., KAWAI, S., HISHIDA, A., WAKAI, K., HAMAJIMA, N., ICHIDA, K., SAKURAI, Y., KATO, Y., SHIMIZU, T. & SHINOMIYA, N. 2013. Common missense variant of monocarboxylate transporter 9 (MCT9/SLC16A9) gene is associated with renal overload gout, but not with all gout susceptibility. *Hum Cell,* 26**,** 133-6.

NAUKKARINEN, J., HEINONEN, S., HAKKARAINEN, A., LUNDBOM, J., VUOLTEENAHO, K., SAARINEN, L., HAUTANIEMI, S., RODRIGUEZ, A., FRUHBECK, G., PAJUNEN, P., HYOTYLAINEN, T., ORESIC, M., MOILANEN, E., SUOMALAINEN, A., LUNDBOM, N., KAPRIO, J., RISSANEN, A. & PIETILAINEN, K. H. 2014. Characterising metabolically healthy obesity in weight-discordant monozygotic twins. *Diabetologia,* 57**,** 167-76.

NEALE, M. C., BOKER, S. M., XIE, G. & MAES, H. 2003. Mx: Statistical modeling. Richmond: Department of Psychiatry, Medical College of Virginia.

NEALE, M. C., CARDON, L. R. & ORGANIZATION, N. A. T. 1992. *Methodology for genetic studies of twins and families,* Dordrecht, Kluwer Academic Publishers.

NETTLETON, J. A., FOLLIS, J. L., NGWA, J. S., SMITH, C. E., AHMAD, S., TANAKA, T., WOJCZYNSKI, M. K., VOORTMAN, T., LEMAITRE, R. N., KRISTIANSSON, K., NUOTIO, M. L., HOUSTON, D. K., PERALA, M. M., QI, Q., SONESTEDT, E., MANICHAIKUL, A., KANONI, S., GANNA, A., MIKKILA, V., NORTH, K. E., SISCOVICK, D. S., HARALD, K., MCKEOWN, N. M., JOHANSSON, I., RISSANEN, H., LIU, Y., LAHTI, J., HU, F. B., BANDINELLI, S., RUKH, G., RICH, S., BOOIJ, L., DMITRIOU, M., AX, E., RAITAKARI, O., MUKAMAL, K., MANNISTO, S., HALLMANS, G., JULA, A., ERICSON, U., JACOBS, D. R., JR., VAN ROOIJ, F. J., DELOUKAS, P., SJOGREN, P., KAHONEN, M., DJOUSSE, L., PEROLA, M., BARROSO, I., HOFMAN, A., STIRRUPS, K., VIIKARI, J., UITTERLINDEN, A. G., KALAFATI, I. P., FRANCO, O. H., MOZAFFARIAN, D., SALOMAA, V., BORECKI, I. B., KNEKT, P., KRITCHEVSKY, S. B., ERIKSSON, J. G., DEDOUSSIS, G. V., QI, L., FERRUCCI, L., ORHO-MELANDER, M., ZILLIKENS, M. C., INGELSSON, E., LEHTIMAKI, T., RENSTROM, F., CUPPLES, L. A., LOOS, R. J. & FRANKS, P. W. 2015. Gene x dietary pattern

interactions in obesity: analysis of up to 68 317 adults of European ancestry. *Hum Mol Genet,* 24**,** 4728-38.
NEWGARD, C. B., AN, J., BAIN, J. R., MUEHLBAUER, M. J., STEVENS, R. D., LIEN, L. F., HAQQ, A. M., SHAH, S. H., ARLOTTO, M., SLENTZ, C. A., ROCHON, J., GALLUP, D., ILKAYEVA, O., WENNER, B. R., YANCY, W. S., JR., EISENSON, H., MUSANTE, G., SURWIT, R. S., MILLINGTON, D. S., BUTLER, M. D. & SVETKEY, L. P. 2009. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab,* 9**,** 311-26.
NEYRINCK, A. M., VAN HEE, V. F., BINDELS, L. B., DE BACKER, F., CANI, P. D. & DELZENNE, N. M. 2013. Polyphenol-rich extract of pomegranate peel alleviates tissue inflammation and hypercholesterolaemia in high-fat diet-induced obese mice: potential implication of the gut microbiota. *Br J Nutr,* 109**,** 802-9.
NICHOLSON, J. K., HOLMES, E., KINROSS, J., BURCELIN, R., GIBSON, G., JIA, W. & PETTERSSON, S. 2012. Host-gut microbiota metabolic interactions. *Science,* 336**,** 1262-7.
NORDESTGAARD, A. T., THOMSEN, M. & NORDESTGAARD, B. G. 2015. Coffee intake and risk of obesity, metabolic syndrome and type 2 diabetes: a Mendelian randomization study. *Int J Epidemiol,* 44**,** 551-65.
NORHEIM, F., GJELSTAD, I. M., HJORTH, M., VINKNES, K. J., LANGLEITE, T. M., HOLEN, T., JENSEN, J., DALEN, K. T., KARLSEN, A. S., KIELLAND, A., RUSTAN, A. C. & DREVON, C. A. 2012. Molecular nutrition research: the modern way of performing nutritional science. *Nutrients,* 4**,** 1898-944.
O'KEEFE, S. J., LI, J. V. & LAHTI, L. 2015. Fat, fibre and cancer risk in African Americans and rural Africans. 6**,** 6342.
O'SULLIVAN, A., GIBNEY, M. J. & BRENNAN, L. 2011. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *Am J Clin Nutr,* 93**,** 314-21.
ODEGAARD, A. O., CHOH, A. C., CZERWINSKI, S. A., TOWNE, B. & DEMERATH, E. W. 2012. Sugar-sweetened and diet beverages in relation to visceral adipose tissue. *Obesity (Silver Spring),* 20**,** 689-91.
OHARA, H., MATSUMOTO, H., ITO, K., IWAI, K. & SATO, K. 2007. Comparison of quantity and structures of hydroxyproline-containing peptides in human blood after oral ingestion of gelatin hydrolysates from different sources. *J Agric Food Chem,* 55**,** 1532-5.
OLZA, J., AGUILERA, C. M., GIL-CAMPOS, M., LEIS, R., BUENO, G., VALLE, M., CANETE, R., TOJO, R., MORENO, L. A. & GIL, A. 2015. A Continuous Metabolic Syndrome Score Is Associated with Specific Biomarkers of Inflammation and CVD Risk in Prepubertal Children. *Ann Nutr Metab,* 66**,** 72-9.
OTT, S. J., MUSFELDT, M., WENDEROTH, D. F., HAMPE, J., BRANT, O., FOLSCH, U. R., TIMMIS, K. N. & SCHREIBER, S. 2004. Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut,* 53**,** 685-93.
OU, K., SARNOSKI, P., SCHNEIDER, K. R., SONG, K., KHOO, C. & GU, L. 2014. Microbial catabolism of procyanidins by human gut microbiota. *Mol Nutr Food Res,* 58**,** 2196-205.
PALKA, J. A. 1996. The role of prolidase as an enzyme participating in the metabolism of collagen. *Rocz Akad Med Bialymst,* 41**,** 149-60.
PALLISTER, T., JENNINGS, A., MOHNEY, R. P., YARAND, D., MANGINO, M., CASSIDY, A., MACGREGOR, A., SPECTOR, T. D. & MENNI, C. 2016. Characterizing Blood Metabolomics Profiles Associated with Self-Reported Food Intakes in Female Twins. *PLoS One,* 11**,** e0158568.
PALLISTER, T., SHARAFI, M., LACHANCE, G., PIRASTU, N., MOHNEY, R. P., MACGREGOR, A., FESKENS, E. J., DUFFY, V., SPECTOR, T. D. & MENNI, C. 2015. Food Preference Patterns in a UK Twin Cohort. *Twin Res Hum Genet,* 18**,** 793-805.
PALLISTER, T., SPECTOR, T. D. & MENNI, C. 2014. Twin studies advance the understanding of gene-environment interplay in human nutrigenomics. *Nutr Res Rev,* 27**,** 242-51.
PATTI, G. J., YANES, O. & SIUZDAK, G. 2012. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol,* 13**,** 263-9.
PERREAULT, M., BIALEK, A., TROTTIER, J., VERREAULT, M., CARON, P., MILKIEWICZ, P. & BARBIER, O. 2013. Role of glucuronidation for hepatic detoxification and urinary elimination of toxic bile acids during biliary obstruction. *PLoS One,* 8**,** e80994.
PETERSEN, A. K., ZEILINGER, S., KASTENMULLER, G., ROMISCH-MARGL, W., BRUGGER, M., PETERS, A., MEISINGER, C., STRAUCH, K., HENGSTENBERG, C., PAGEL, P., HUBER, F., MOHNEY, R. P., GRALLERT, H., ILLIG, T., ADAMSKI, J., WALDENBERGER, M., GIEGER, C. & SUHRE, K. 2014. Epigenetics meets

metabolomics: an epigenome-wide association study with blood serum metabolic traits. *Hum Mol Genet,* 23**,** 534-45.

PIETILAINEN, K. H., SYSI-AHO, M., RISSANEN, A., SEPPANEN-LAAKSO, T., YKI-JARVINEN, H., KAPRIO, J. & ORESIC, M. 2007. Acquired obesity is associated with changes in the serum lipidomic profile independent of genetic effects--a monozygotic twin study. *PLoS One,* 2**,** e218.

PIMPIN, L., AMBROSINI, G. L., LLEWELLYN, C. H., JOHNSON, L., VAN JAARSVELD, C. H., JEBB, S. A. & WARDLE, J. 2013. Dietary intake of young twins: nature or nurture? *Am J Clin Nutr,* 98**,** 1326-34.

POLLARD, S. L., ZACHARY, D. A., WINGERT, K., BOOKER, S. S. & SURKAN, P. J. 2014. Family and Community Influences on Diabetes-Related Dietary Change in a Low-Income Urban Neighborhood. *Diabetes Educ,* 40**,** 462-469.

QI, Q., DURST, R., SCHWARZFUCHS, D., LEITERSDORF, E., SHPITZEN, S., LI, Y., WU, H., CHAMPAGNE, C. M., HU, F. B., STAMPFER, M. J., BRAY, G. A., SACKS, F. M., SHAI, I. & QI, L. 2015. CETP genotype and changes in lipid levels in response to weight-loss diet intervention in the POUNDS LOST and DIRECT randomized trials. *J Lipid Res,* 56**,** 713-21.

QIN, J., LI, R., RAES, J., ARUMUGAM, M., BURGDORF, K. S., MANICHANH, C., NIELSEN, T., PONS, N., LEVENEZ, F., YAMADA, T., MENDE, D. R., LI, J., XU, J., LI, S., LI, D., CAO, J., WANG, B., LIANG, H., ZHENG, H., XIE, Y., TAP, J., LEPAGE, P., BERTALAN, M., BATTO, J. M., HANSEN, T., LE PASLIER, D., LINNEBERG, A., NIELSEN, H. B., PELLETIER, E., RENAULT, P., SICHERITZ-PONTEN, T., TURNER, K., ZHU, H., YU, C., LI, S., JIAN, M., ZHOU, Y., LI, Y., ZHANG, X., LI, S., QIN, N., YANG, H., WANG, J., BRUNAK, S., DORE, J., GUARNER, F., KRISTIANSEN, K., PEDERSEN, O., PARKHILL, J., WEISSENBACH, J., BORK, P., EHRLICH, S. D. & WANG, J. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature,* 464**,** 59-65.

QUILES, J. L., HUERTAS, J. R., BATTINO, M., RAMIREZ-TORTOSA, M. C., CASSINELLO, M., MATAIX, J., LOPEZ-FRIAS, M. & MANAS, M. 2002. The intake of fried virgin olive or sunflower oils differentially induces oxidative stress in rat liver microsomes. *Br J Nutr,* 88**,** 57-65.

RABE-HESKETH, S. & SKRONDAL, A. 2008. Classical latent variable models for medical research. *Stat Methods Med Res,* 17**,** 5-32.

RANKINEN, T. & BOUCHARD, C. 2006. Genetics of food intake and eating behavior phenotypes in humans. *Annu Rev Nutr,* 26**,** 413-34.

REVELLES, O., ESPINOSA-URGEL, M., FUHRER, T., SAUER, U. & RAMOS, J. L. 2005. Multiple and interconnected pathways for L-lysine catabolism in Pseudomonas putida KT2440. *J Bacteriol,* 187**,** 7500-10.

REZZI, S., RAMADAN, Z., MARTIN, F. P., FAY, L. B., VAN BLADEREN, P., LINDON, J. C., NICHOLSON, J. K. & KOCHHAR, S. 2007. Human metabolic phenotypes link directly to specific dietary preferences in healthy individuals. *J Proteome Res,* 6**,** 4469-77.

RICHARDS, J. B., RIVADENEIRA, F., INOUYE, M., PASTINEN, T. M., SORANZO, N., WILSON, S. G., ANDREW, T., FALCHI, M., GWILLIAM, R., AHMADI, K. R., VALDES, A. M., ARP, P., WHITTAKER, P., VERLAAN, D. J., JHAMAI, M., KUMANDURI, V., MOORHOUSE, M., VAN MEURS, J. B., HOFMAN, A., POLS, H. A., HART, D., ZHAI, G., KATO, B. S., MULLIN, B. H., ZHANG, F., DELOUKAS, P., UITTERLINDEN, A. G. & SPECTOR, T. D. 2008. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet,* 371**,** 1505-12.

RICHTER, B., KOLLER, L., HOHENSINNER, P. J., ZORN, G., BREKALO, M., BERGER, R., MORTL, D., MAURER, G., PACHER, R., HUBER, K., WOJTA, J., HULSMANN, M. & NIESSNER, A. 2013. A multi-biomarker risk score improves prediction of long-term mortality in patients with advanced heart failure. *Int J Cardiol,* 168**,** 1251-7.

RIDAURA, V. K., FAITH, J. J., REY, F. E., CHENG, J., DUNCAN, A. E., KAU, A. L., GRIFFIN, N. W., LOMBARD, V., HENRISSAT, B., BAIN, J. R., MUEHLBAUER, M. J., ILKAYEVA, O., SEMENKOVICH, C. F., FUNAI, K., HAYASHI, D. K., LYLE, B. J., MARTINI, M. C., URSELL, L. K., CLEMENTE, J. C., VAN TREUREN, W., WALTERS, W. A., KNIGHT, R., NEWGARD, C. B., HEATH, A. C. & GORDON, J. I. 2013. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science,* 341**,** 1241214.

RODRIGUEZ, A., COSTA-BAUZA, A., SAEZ-TORRES, C., RODRIGO, D. & GRASES, F. 2015. HPLC method for urinary theobromine determination: Effect of consumption of cocoa products on theobromine urinary excretion in children. *Clin Biochem*.

RODRIGUEZ-SUAREZ, E., MATO, J. M. & ELORTZA, F. 2012. Proteomics analysis of human nonalcoholic fatty liver. *Methods Mol Biol,* 909**,** 241-58.

ROMAGUERA, D., ANGQUIST, L., DU, H., JAKOBSEN, M. U., FOROUHI, N. G., HALKJAER, J., FESKENS, E. J., VAN DER, A. D., MASALA, G., STEFFEN, A., PALLI, D., WAREHAM, N. J., OVERVAD, K., TJONNELAND, A., BOEING, H., RIBOLI, E. & SORENSEN, T. I. 2010. Dietary determinants of changes in waist circumference adjusted for body mass index - a proxy measure of visceral adiposity. *PLoS One,* 5**,** e11588.

ROMAGUERA, D., ANGQUIST, L., DU, H., JAKOBSEN, M. U., FOROUHI, N. G., HALKJAER, J., FESKENS, E. J., VAN DER, A. D., MASALA, G., STEFFEN, A., PALLI, D., WAREHAM, N. J., OVERVAD, K., TJONNELAND, A., BOEING, H., RIBOLI, E. & SORENSEN, T. I. 2011. Food composition of the diet in relation to changes in waist circumference adjusted for body mass index. *PLoS One,* 6**,** e23384.

RÖMISCH-MARGL, W., PREHN, C., BOGUMIL, R., RÖHRING, C., SUHRE, K. & ADAMSKI, J. 2012. Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. *Metabolomics,* 8**,** 133-142.

ROOPCHAND, D. E., CARMODY, R. N., KUHN, P., MOSKAL, K., ROJAS-SILVA, P., TURNBAUGH, P. J. & RASKIN, I. 2015. Dietary Polyphenols Promote Growth of the Gut Bacterium Akkermansia muciniphila and Attenuate High-Fat Diet-Induced Metabolic Syndrome. *Diabetes,* 64**,** 2847-58.

SACQUET, E., PARQUET, M., RIOTTOT, M., RAIZMAN, A., JARRIGE, P., HUGUET, C. & INFANTE, R. 1983. Intestinal absorption, excretion, and biotransformation of hyodeoxycholic acid in man. *J Lipid Res,* 24**,** 604-13.

SAHYOUN, N. R., JACQUES, P. F., ZHANG, X. L., JUAN, W. & MCKEOWN, N. M. 2006. Whole-grain intake is inversely associated with the metabolic syndrome and mortality in older adults. *Am J Clin Nutr,* 83**,** 124-31.

SAMPSON, J. N., BOCA, S. M., SHU, X. O., STOLZENBERG-SOLOMON, R. Z., MATTHEWS, C. E., HSING, A. W., TAN, Y. T., JI, B. T., CHOW, W. H., CAI, Q., LIU DA, K., YANG, G., XIANG, Y. B., ZHENG, W., SINHA, R., CROSS, A. J. & MOORE, S. C. 2013. Metabolomics in epidemiology: sources of variability in metabolite measurements and implications. *Cancer Epidemiol Biomarkers Prev,* 22**,** 631-40.

SANTAREN, I. D., WATKINS, S. M., LIESE, A. D., WAGENKNECHT, L. E., REWERS, M. J., HAFFNER, S. M., LORENZO, C. & HANLEY, A. J. 2014. Serum pentadecanoic acid (15:0), a short-term marker of dairy food intake, is inversely associated with incident type 2 diabetes and its underlying disorders. *Am J Clin Nutr,* 100**,** 1532-40.

SCALBERT, A., BRENNAN, L., MANACH, C., ANDRES-LACUEVA, C., DRAGSTED, L. O., DRAPER, J., RAPPAPORT, S. M., VAN DER HOOFT, J. J. & WISHART, D. S. 2014. The food metabolome: a window over dietary exposure. *Am J Clin Nutr,* 99**,** 1286-308.

SCARR, S. 1968. Environmental bias in twin studies. *Eugen Q,* 15**,** 34-40.

SCHEIBEHENNE, B., TODD, P. M., VAN DEN BERG, S. M., HATEMI, P. K., EAVES, L. J. & VOGLER, C. 2014. Genetic influences on dietary variety - results from a twin study. *Appetite,* 77**,** 131-8.

SCHER, J. U., UBEDA, C., ARTACHO, A., ATTUR, M., ISAAC, S., REDDY, S. M., MARMON, S., NEIMANN, A., BRUSCA, S., PATEL, T., MANASSON, J., PAMER, E. G., LITTMAN, D. R. & ABRAMSON, S. B. 2015. Decreased bacterial diversity characterizes the altered gut microbiota in patients with psoriatic arthritis, resembling dysbiosis in inflammatory bowel disease. *Arthritis Rheumatol,* 67**,** 128-39.

SELLICK, C. A., KNIGHT, D., CROXFORD, A. S., MAQSOOD, A. R., STEPHENS, G. M., GOODACRE, R. & DICKSON, A. J. 2010. Evaluation of extraction processes for intracellular metabolite profiling of mammalian cells: matching extraction approaches to cell type and metabolite targets. *Metabolomics,* 6**,** 427-438.

SERINO, M., LUCHE, E., GRES, S., BAYLAC, A., BERGE, M., CENAC, C., WAGET, A., KLOPP, P., IACOVONI, J., KLOPP, C., MARIETTE, J., BOUCHEZ, O., LLUCH, J., OUARNE, F., MONSAN, P., VALET, P., ROQUES, C., AMAR, J., BOULOUMIE, A., THEODOROU, V. & BURCELIN, R. 2012. Metabolic adaptation to a high-fat diet is associated with a change in the gut microbiota. *Gut,* 61**,** 543-53.

SHEARER, J., DUGGAN, G., WELJIE, A., HITTEL, D. S., WASSERMAN, D. H. & VOGEL, H. J. 2008. Metabolomic profiling of dietary-induced insulin resistance in the high fat-fed C57BL/6J mouse. *Diabetes Obes Metab,* 10**,** 950-8.

SHEFLIN, A. M., BORRESEN, E. C., WDOWIK, M. J., RAO, S., BROWN, R. J., HEUBERGER, A. L., BROECKLING, C. D., WEIR, T. L. & RYAN, E. P. 2015. Pilot dietary intervention with heat-stabilized rice bran modulates stool microbiota and metabolites in healthy adults. *Nutrients,* 7**,** 1282-300.

SHEN, W., WOLF, P. G., CARBONERO, F., ZHONG, W., REID, T., GASKINS, H. R. & MCINTOSH, M. K. 2014. Intestinal and systemic inflammatory responses are positively

associated with sulfidogenic bacteria abundance in high-fat-fed male C57BL/6J mice. *J Nutr,* 144**,** 1181-7.
SHIM, J. S., OH, K. & KIM, H. C. 2014. Dietary assessment methods in epidemiologic studies. *Epidemiol Health,* 36**,** e2014009.
SHIN, S. Y., FAUMAN, E. B., PETERSEN, A. K., KRUMSIEK, J., SANTOS, R., HUANG, J., ARNOLD, M., ERTE, I., FORGETTA, V., YANG, T. P., WALTER, K., MENNI, C., CHEN, L., VASQUEZ, L., VALDES, A. M., HYDE, C. L., WANG, V., ZIEMEK, D., ROBERTS, P., XI, L., GRUNDBERG, E., WALDENBERGER, M., RICHARDS, J. B., MOHNEY, R. P., MILBURN, M. V., JOHN, S. L., TRIMMER, J. & THEIS, F. J. 2014. An atlas of genetic influences on human blood metabolites. *Nat Genet,* 46**,** 543-50.
SHOAIE, S., GHAFFARI, P., KOVATCHEVA-DATCHARY, P., MARDINOGLU, A., SEN, P., PUJOS-GUILLOT, E., DE WOUTERS, T., JUSTE, C., RIZKALLA, S., CHILLOUX, J., HOYLES, L., NICHOLSON, J. K., DORE, J., DUMAS, M. E., CLEMENT, K., BACKHED, F. & NIELSEN, J. 2015. Quantifying Diet-Induced Metabolic Changes of the Human Gut Microbiome. *Cell Metab,* 22**,** 320-31.
SIMOES, C. D., MAUKONEN, J., KAPRIO, J., RISSANEN, A., PIETILAINEN, K. H. & SAARELA, M. 2013. Habitual dietary intake is associated with stool microbiota composition in monozygotic twins. *J Nutr,* 143**,** 417-23.
SMITH, C. E., FOLLIS, J. L., NETTLETON, J. A., FOY, M., WU, J. H., MA, Y., TANAKA, T., MANICHAKUL, A. W., WU, H., CHU, A. Y., STEFFEN, L. M., FORNAGE, M., MOZAFFARIAN, D., KABAGAMBE, E. K., FERRUCI, L., CHEN, Y. D., RICH, S. S., DJOUSSE, L., RIDKER, P. M., TANG, W., MCKNIGHT, B., TSAI, M. Y., BANDINELLI, S., ROTTER, J. I., HU, F. B., CHASMAN, D. I., PSATY, B. M., ARNETT, D. K., KING, I. B., SUN, Q., WANG, L., LUMLEY, T., CHIUVE, S. E., SISCOVICK, D. S., ORDOVAS, J. M. & LEMAITRE, R. N. 2015. Dietary fatty acids modulate associations between genetic variants and circulating fatty acids in plasma and erythrocyte membranes: Meta-analysis of nine studies in the CHARGE consortium. *Mol Nutr Food Res,* 59**,** 1373-83.
SMITH, E. A. & MACFARLANE, G. T. 1996. Enumeration of human colonic bacteria producing phenolic and indolic compounds: effects of pH, carbohydrate availability and retention time on dissimilatory aromatic amino acid metabolism. *J Appl Bacteriol,* 81**,** 288-302.
SOKOL, H., PIGNEUR, B., WATTERLOT, L., LAKHDARI, O., BERMUDEZ-HUMARAN, L. G., GRATADOUX, J. J., BLUGEON, S., BRIDONNEAU, C., FURET, J. P., CORTHIER, G., GRANGETTE, C., VASQUEZ, N., POCHART, P., TRUGNAN, G., THOMAS, G., BLOTTIERE, H. M., DORE, J., MARTEAU, P., SEKSIK, P. & LANGELLA, P. 2008. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A,* 105**,** 16731-6.
SONG, J., KWON, N., LEE, M. H., KO, Y. G., LEE, J. H. & KIM, O. Y. 2014. Association of serum phospholipid PUFAs with cardiometabolic risk: beneficial effect of DHA on the suppression of vascular proliferation/inflammation. *Clin Biochem,* 47**,** 361-8.
SONG, M. Y., WANG, J. H., EOM, T. & KIM, H. 2015. Schisandra chinensis fruit modulates the gut microbiota composition in association with metabolic markers in obese women: a randomized, double-blind placebo-controlled study. *Nutr Res,* 35**,** 655-63.
SONNENBURG, E. D., SMITS, S. A., TIKHONOV, M., HIGGINBOTTOM, S. K., WINGREEN, N. S. & SONNENBURG, J. L. 2016. Diet-induced extinctions in the gut microbiota compound over generations. *Nature,* 529**,** 212-5.
SPECTOR, T. D. & WILLIAMS, F. M. 2006. The UK Adult Twin Registry (TwinsUK). *Twin Res Hum Genet,* 9**,** 899-906.
SPENCER, M. D., HAMP, T. J., REID, R. W., FISCHER, L. M., ZEISEL, S. H. & FODOR, A. A. 2011. Association between composition of the human gastrointestinal microbiome and development of fatty liver with choline deficiency. *Gastroenterology,* 140**,** 976-86.
ST-PIERRE, J., LEMIEUX, I., VOHL, M. C., PERRON, P., TREMBLAY, G., DESPRES, J. P. & GAUDET, D. 2002. Contribution of abdominal obesity and hypertriglyceridemia to impaired fasting glucose and coronary artery disease. *Am J Cardiol,* 90**,** 15-8.
SUHRE, K., SHIN, S. Y., PETERSEN, A. K., MOHNEY, R. P., MEREDITH, D., WAGELE, B., ALTMAIER, E., DELOUKAS, P., ERDMANN, J., GRUNDBERG, E., HAMMOND, C. J., DE ANGELIS, M. H., KASTENMULLER, G., KOTTGEN, A., KRONENBERG, F., MANGINO, M., MEISINGER, C., MEITINGER, T., MEWES, H. W., MILBURN, M. V., PREHN, C., RAFFLER, J., RIED, J. S., ROMISCH-MARGL, W., SAMANI, N. J., SMALL, K. S., WICHMANN, H. E., ZHAI, G., ILLIG, T., SPECTOR, T. D., ADAMSKI, J., SORANZO, N. & GIEGER, C. 2011. Human metabolic individuality in biomedical and pharmaceutical research. *Nature,* 477**,** 54-60.

SUN, H. & LIU, D. 2015. Hydrodynamic delivery of interleukin 15 gene promotes resistance to high fat diet-induced obesity, fatty liver and improves glucose homeostasis. *Gene Ther,* 22**,** 341-7.

SUZUKI, T. & TAKAHASHI, E. 1975. Biosynthesis of caffeine by tea-leaf extracts. Enzymic formation of theobromine from 7-methylxanthine and of caffeine from theobromine. *Biochem J,* 146**,** 87-96.

TAIRA, R., YAMAGUCHI, S., SHIMIZU, K., NAKAMURA, K., AYABE, T. & TAIRA, T. 2015. Bacterial cell wall components regulate adipokine secretion from visceral adipocytes. *J Clin Biochem Nutr,* 56**,** 149-54.

TAP, J., FURET, J. P., BENSAADA, M., PHILIPPE, C., ROTH, H., RABOT, S., LAKHDARI, O., LOMBARD, V., HENRISSAT, B., CORTHIER, G., FONTAINE, E., DORE, J. & LECLERC, M. 2015. Gut microbiota richness promotes its stability upon increased dietary fibre intake in healthy adults. *Environ Microbiol,* 17**,** 4954-64.

TEMELLINI, A., MOGAVERO, S., GIULIANOTTI, P. C., PIETRABISSA, A., MOSCA, F. & PACIFICI, G. M. 1993. Conjugation of benzoic acid with glycine in human liver and kidney: a study on the interindividual variability. *Xenobiotica,* 23**,** 1427-33.

TEO, Y. Y., INOUYE, M., SMALL, K. S., GWILLIAM, R., DELOUKAS, P., KWIATKOWSKI, D. P. & CLARK, T. G. 2007. A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics,* 23**,** 2741-6.

TEUCHER, B., SKINNER, J., SKIDMORE, P. M., CASSIDY, A., FAIRWEATHER-TAIT, S. J., HOOPER, L., ROE, M. A., FOXALL, R., OYSTON, S. L., CHERKAS, L. F., PERKS, U. C., SPECTOR, T. D. & MACGREGOR, A. J. 2007. Dietary patterns and heritability of food choice in a UK female twin cohort. *Twin Res Hum Genet,* 10**,** 734-48.

TIMS, S., DEROM, C., JONKERS, D. M., VLIETINCK, R., SARIS, W. H., KLEEREBEZEM, M., DE VOS, W. M. & ZOETENDAL, E. G. 2013. Microbiota conservation and BMI signatures in adult monozygotic twins. *Isme j,* 7**,** 707-17.

TORRIS, C., MOLIN, M. & CVANCAROVA SMASTUEN, M. 2014. Fish consumption and its possible preventive role on the development and prevalence of metabolic syndrome - a systematic review. *Diabetol Metab Syndr,* 6**,** 112.

TOUSOULIS, D., PLASTIRAS, A., SIASOS, G., OIKONOMOU, E., VERVENIOTIS, A., KOKKOU, E., MANIATIS, K., GOULIOPOULOS, N., MILIOU, A., PARASKEVOPOULOS, T. & STEFANADIS, C. 2014. Omega-3 PUFAs improved endothelial function and arterial stiffness with a parallel antiinflammatory effect in adults with metabolic syndrome. *Atherosclerosis,* 232**,** 10-6.

TRICHOPOULOU, A., COSTACOU, T., BAMIA, C. & TRICHOPOULOS, D. 2003. Adherence to a Mediterranean diet and survival in a Greek population. *N Engl J Med,* 348**,** 2599-608.

TRICHOPOULOU, A., ORFANOS, P., NORAT, T., BUENO-DE-MESQUITA, B., OCKE, M. C., PEETERS, P. H., VAN DER SCHOUW, Y. T., BOEING, H., HOFFMANN, K., BOFFETTA, P., NAGEL, G., MASALA, G., KROGH, V., PANICO, S., TUMINO, R., VINEIS, P., BAMIA, C., NASKA, A., BENETOU, V., FERRARI, P., SLIMANI, N., PERA, G., MARTINEZ-GARCIA, C., NAVARRO, C., RODRIGUEZ-BARRANCO, M., DORRONSORO, M., SPENCER, E. A., KEY, T. J., BINGHAM, S., KHAW, K. T., KESSE, E., CLAVEL-CHAPELON, F., BOUTRON-RUAULT, M. C., BERGLUND, G., WIRFALT, E., HALLMANS, G., JOHANSSON, I., TJONNELAND, A., OLSEN, A., OVERVAD, K., HUNDBORG, H. H., RIBOLI, E. & TRICHOPOULOS, D. 2005. Modified Mediterranean diet and survival: EPIC-elderly prospective cohort study. *Bmj,* 330**,** 991.

TUCKER, K. L., SMITH, C. E., LAI, C. Q. & ORDOVAS, J. M. 2013. Quantifying diet for nutrigenomic studies. *Annu Rev Nutr,* 33**,** 349-71.

TURNBAUGH, P. J., BACKHED, F., FULTON, L. & GORDON, J. I. 2008. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe,* 3**,** 213-23.

TURNBAUGH, P. J., HAMADY, M., YATSUNENKO, T., CANTAREL, B. L., DUNCAN, A., LEY, R. E., SOGIN, M. L., JONES, W. J., ROE, B. A., AFFOURTIT, J. P., EGHOLM, M., HENRISSAT, B., HEATH, A. C., KNIGHT, R. & GORDON, J. I. 2009. A core gut microbiome in obese and lean twins. *Nature,* 457**,** 480-4.

URSELL, L. K., HAISER, H. J., VAN TREUREN, W., GARG, N., REDDIVARI, L., VANAMALA, J., DORRESTEIN, P. C., TURNBAUGH, P. J. & KNIGHT, R. 2014. The intestinal metabolome: an intersection between microbiota and host. *Gastroenterology,* 146**,** 1470-6.

VAN DEN BERG, L., HENNEMAN, P., WILLEMS VAN DIJK, K., DELEMARRE-VAN DE WAAL, H. A., OOSTRA, B. A., VAN DUIJN, C. M. & JANSSENS, A. C. 2013. Heritability of dietary food intake patterns. *Acta Diabetol,* 50**,** 721-6.

VAN DEN BREE, M. B., EAVES, L. J. & DWYER, J. T. 1999. Genetic and environmental influences on eating patterns of twins aged >/=50 y. *Am J Clin Nutr,* 70**,** 456-65.
VAN DONGEN, J., SLAGBOOM, P. E., DRAISMA, H. H., MARTIN, N. G. & BOOMSMA, D. I. 2012. The continuing value of twin studies in the omics era. *Nat Rev Genet,* 13**,** 640-53.
VAN DORSTEN, F. A., DAYKIN, C. A., MULDER, T. P. & VAN DUYNHOVEN, J. P. 2006. Metabonomics approach to determine metabolic differences between green tea and black tea consumption. *J Agric Food Chem,* 54**,** 6929-38.
VAN OMMEN, B., BOUWMAN, J., DRAGSTED, L. O., DREVON, C. A., ELLIOTT, R., DE GROOT, P., KAPUT, J., MATHERS, J. C., MULLER, M., PEPPING, F., SAITO, J., SCALBERT, A., RADONJIC, M., ROCCA-SERRA, P., TRAVIS, A., WOPEREIS, S. & EVELO, C. T. 2010. Challenges of molecular nutrition research 6: the nutritional phenotype database to store, share and evaluate nutritional systems biology studies. *Genes Nutr,* 5**,** 189-203.
VAN OMMEN, B. & STIERUM, R. 2002. Nutrigenomics: exploiting systems biology in the nutrition and health arena. *Curr Opin Biotechnol,* 13**,** 517-21.
VARVEL, S. A., POTTALA, J. V., THISELTON, D. L., CAFFREY, R., DALL, T., SASINOWSKI, M., MCCONNELL, J. P., WARNICK, G. R., VOROS, S. & GRAHAM, T. E. 2014. Serum alpha-hydroxybutyrate (alpha-HB) predicts elevated 1 h glucose levels and early-phase beta-cell dysfunction during OGTT. *BMJ Open Diabetes Res Care,* 2**,** e000038.
VEENSTRA, T. D. 2012. Metabolomics: the final frontier? *Genome Med,* 4**,** 40.
VERNOCCHI, P., VANNINI, L., GOTTARDI, D., DEL CHIERICO, F., SERRAZANETTI, D. I., NDAGIJIMANA, M. & GUERZONI, M. E. 2012. Integration of datasets from different analytical techniques to assess the impact of nutrition on human metabolome. *Front Cell Infect Microbiol,* 2**,** 156.
VISSCHER, P. M., HILL, W. G. & WRAY, N. R. 2008. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet,* 9**,** 255-66.
VITAL, M., HOWE, A. C. & TIEDJE, J. M. 2014. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *MBio,* 5**,** e00889.
WADE, J., MILNER, J. & KRONDL, M. 1981. Evidence for a physiological regulation of food selection and nutrient intake in twins. *Am J Clin Nutr,* 34**,** 143-7.
WALDRAM, A., HOLMES, E., WANG, Y., RANTALAINEN, M., WILSON, I. D., TUOHY, K. M., MCCARTNEY, A. L., GIBSON, G. R. & NICHOLSON, J. K. 2009. Top-down systems biology modeling of host metabotype-microbiome associations in obese rodents. *J Proteome Res,* 8**,** 2361-75.
WALKER, A. W., INCE, J., DUNCAN, S. H., WEBSTER, L. M., HOLTROP, G., ZE, X., BROWN, D., STARES, M. D., SCOTT, P., BERGERAT, A., LOUIS, P., MCINTOSH, F., JOHNSTONE, A. M., LOBLEY, G. E., PARKHILL, J. & FLINT, H. J. 2011. Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *Isme j,* 5**,** 220-30.
WALLNER, S. & SCHMITZ, G. 2011. Plasmalogens the neglected regulatory and scavenging lipid species. *Chem Phys Lipids,* 164**,** 573-89.
WALRADT, J. P. P., A. O.; KINLIN, T.; MURALIDHARA, R.; SANDERSON, A. 1971. Volatile components of roasted peanuts. *Journal of Agricultural and Food Chemistry,* 19**,** 972-979.
WALSH, M. C., BRENNAN, L., PUJOS-GUILLOT, E., SEBEDIO, J. L., SCALBERT, A., FAGAN, A., HIGGINS, D. G. & GIBNEY, M. J. 2007. Influence of acute phytochemical intake on human urinary metabolomic profiles. *Am J Clin Nutr,* 86**,** 1687-93.
WANG, Z., KLIPFELL, E., BENNETT, B. J., KOETH, R., LEVISON, B. S., DUGAR, B., FELDSTEIN, A. E., BRITT, E. B., FU, X., CHUNG, Y. M., WU, Y., SCHAUER, P., SMITH, J. D., ALLAYEE, H., TANG, W. H., DIDONATO, J. A., LUSIS, A. J. & HAZEN, S. L. 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature,* 472**,** 57-63.
WARDLE, J., SANDERSON, S., LEIGH GIBSON, E. & RAPOPORT, L. 2001. Factor-analytic structure of food preferences in four-year-old children in the UK. *Appetite,* 37**,** 217-23.
WEBER, D., OEFNER, P. J., HIERGEIST, A., KOESTLER, J., GESSNER, A., WEBER, M., HAHN, J., WOLFF, D., STAMMLER, F., SPANG, R., HERR, W., DETTMER, K. & HOLLER, E. 2015. Low urinary indoxyl sulfate levels early after transplantation reflect a disrupted microbiome and are associated with poor outcome. *Blood,* 126**,** 1723-8.
WESTERTERP, K. R. & GORIS, A. H. 2002. Validity of the assessment of dietary intake: problems of misreporting. *Curr Opin Clin Nutr Metab Care,* 5**,** 489-93.
WICHMANN, H. E., GIEGER, C. & ILLIG, T. 2005. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen,* 67 Suppl 1**,** S26-30.

WIKOFF, W. R., ANFORA, A. T., LIU, J., SCHULTZ, P. G., LESLEY, S. A., PETERS, E. C. & SIUZDAK, G. 2009. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc Natl Acad Sci U S A,* 106**,** 3698-703.

WIKOFF, W. R., FRYE, R. F., ZHU, H., GONG, Y., BOYLE, S., CHURCHILL, E., COOPER-DEHOFF, R. M., BEITELSHEES, A. L., CHAPMAN, A. B., FIEHN, O., JOHNSON, J. A. & KADDURAH-DAOUK, R. 2013. Pharmacometabolomics reveals racial differences in response to atenolol treatment. *PLoS One,* 8**,** e57639.

WILLETT, W. & STAMPFER, M. J. 1986. Total energy intake: implications for epidemiologic analyses. *Am J Epidemiol,* 124**,** 17-27.

WILLIAMS, R. E., LENZ, E. M., EVANS, J. A., WILSON, I. D., GRANGER, J. H., PLUMB, R. S. & STUMPF, C. L. 2005. A combined (1)H NMR and HPLC-MS-based metabonomic study of urine from obese (fa/fa) Zucker and normal Wistar-derived rats. *J Pharm Biomed Anal,* 38**,** 465-71.

WINKLER, G. & DORING, A. 1998. Validation of a short qualitative food frequency list used in several German large scale surveys. *Z Ernahrungswiss,* 37**,** 234-41.

WITTENBECHER, C., MUHLENBRUCH, K., KROGER, J., JACOBS, S., KUXHAUS, O., FLOEGEL, A., FRITSCHE, A., PISCHON, T., PREHN, C., ADAMSKI, J., JOOST, H. G., BOEING, H. & SCHULZE, M. B. 2015. Amino acids, lipid metabolites, and ferritin as potential mediators linking red meat consumption to type 2 diabetes. *Am J Clin Nutr,* 101**,** 1241-50.

WOLEVER, T. M., TOSH, S. M., GIBBS, A. L., BRAND-MILLER, J., DUNCAN, A. M., HART, V., LAMARCHE, B., THOMSON, B. A., DUSS, R. & WOOD, P. J. 2010. Physicochemical properties of oat beta-glucan influence its ability to reduce serum LDL cholesterol in humans: a randomized clinical trial. *Am J Clin Nutr,* 92**,** 723-32.

WOTING, A., PFEIFFER, N., LOH, G., KLAUS, S. & BLAUT, M. 2014. Clostridium ramosum promotes high-fat diet-induced obesity in gnotobiotic mouse models. *MBio,* 5**,** e01530-14.

WU, G. D., CHEN, J., HOFFMANN, C., BITTINGER, K., CHEN, Y. Y., KEILBAUGH, S. A., BEWTRA, M., KNIGHTS, D., WALTERS, W. A., KNIGHT, R., SINHA, R., GILROY, E., GUPTA, K., BALDASSANO, R., NESSEL, L., LI, H., BUSHMAN, F. D. & LEWIS, J. D. 2011a. Linking long-term dietary patterns with gut microbial enterotypes. *Science,* 334**,** 105-8.

WU, Y., LI, M., XU, M., BI, Y., LI, X., CHEN, Y., NING, G. & WANG, W. 2011b. Low serum total bilirubin concentrations are associated with increased prevalence of metabolic syndrome in Chinese. *J Diabetes,* 3**,** 217-24.

XU, M., NG, S. S., BRAY, G. A., RYAN, D. H., SACKS, F. M., NING, G. & QI, L. 2015. Dietary Fat Intake Modifies the Effect of a Common Variant in the LIPC Gene on Changes in Serum Lipid Concentrations during a Long-Term Weight-Loss Intervention Trial. *J Nutr,* 145**,** 1289-94.

YAKOOB, M. Y., SHI, P., HU, F. B., CAMPOS, H., REXRODE, K. M., ORAV, E. J., WILLETT, W. C. & MOZAFFARIAN, D. 2014. Circulating biomarkers of dairy fat and risk of incident stroke in U.S. men and women in 2 large prospective cohorts. *Am J Clin Nutr,* 100**,** 1437-47.

YANG, A., PALMER, A. A. & DE WIT, H. 2010a. Genetics of caffeine consumption and responses to caffeine. *Psychopharmacology (Berl),* 211**,** 245-57.

YANG, J., BENYAMIN, B., MCEVOY, B. P., GORDON, S., HENDERS, A. K., NYHOLT, D. R., MADDEN, P. A., HEATH, A. C., MARTIN, N. G., MONTGOMERY, G. W., GODDARD, M. E. & VISSCHER, P. M. 2010b. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet,* 42**,** 565-9.

YANG, L. V., RADU, C. G., WANG, L., RIEDINGER, M. & WITTE, O. N. 2005. Gi-independent macrophage chemotaxis to lysophosphatidylcholine via the immunoregulatory GPCR G2A. *Blood,* 105**,** 1127-34.

YANG, Q., YIN, R. X., CAO, X. L., WU, D. F., CHEN, W. X. & ZHOU, Y. J. 2015. Association of two polymorphisms in the FADS1/FADS2 gene cluster and the risk of coronary artery disease and ischemic stroke. *Int J Clin Exp Pathol,* 8**,** 7318-31.

YATSUNENKO, T., REY, F. E., MANARY, M. J., TREHAN, I., DOMINGUEZ-BELLO, M. G., CONTRERAS, M., MAGRIS, M., HIDALGO, G., BALDASSANO, R. N., ANOKHIN, A. P., HEATH, A. C., WARNER, B., REEDER, J., KUCZYNSKI, J., CAPORASO, J. G., LOZUPONE, C. A., LAUBER, C., CLEMENTE, J. C., KNIGHTS, D., KNIGHT, R. & GORDON, J. I. 2012. Human gut microbiome viewed across age and geography. *Nature,* 486**,** 222-7.

YU, B., ZHENG, Y., ALEXANDER, D., MORRISON, A. C., CORESH, J. & BOERWINKLE, E. 2014. Genetic determinants influencing human serum metabolome among African Americans. *PLoS Genet,* 10**,** e1004212.

ZEISEL, S. H. & DA COSTA, K. A. 2009. Choline: an essential nutrient for public health. *Nutr Rev,* 67**,** 615-23.

ZHENG, Y., YU, B., ALEXANDER, D., STEFFEN, L. M. & BOERWINKLE, E. 2014. Human metabolome associates with dietary intake habits among African Americans in the atherosclerosis risk in communities study. *Am J Epidemiol,* 179**,** 1424-33.

ZIVKOVIC, A. M. & GERMAN, J. B. 2009. Metabolomics for assessment of nutritional status. *Curr Opin Clin Nutr Metab Care,* 12**,** 501-7.

ZULYNIAK, M. A. & MUTCH, D. M. 2011. Harnessing metabolomics for nutrition research. *Curr Pharm Biotechnol,* 12**,** 1005-15.

Appendix A. Chapter 1 Appendices

Table 1. Selection of twin studies estimating heritability of energy and macronutrient intakes (from Pallister et al., 2014).*

						Heritability							
							Macronutrient intake						
							Grams per day			Percent energy			
Author	**Sex**	**Age**	**Dietary assessment method**	**Twin registry**	**Twin pair count**	**Energy intake**	***Fat***	***Protein***	***CHO***	***Fat***	***Protein***	***CHO***
Pimpin *et al.* (2013)[a]	Male + female (*n* = 3605)	21 months	3-day diet diary	Gemini cohort (UK)	384 MZ 832 DZ	12[b]	11	12	9	10	8	9
Liu *et al.* (2013)	Male + female (*n* = 358)	11-13 ($\bar{x}$ = 11.8)	3-day diet diary	USC Twin Study (US)	94 MZ 85 DZ	48	44	31	43			
Faith *et al*. (1999)[b]	Male + female (*n* = 108)	≥ 18	Two Buffet-style meals	New York Obesity Centre twin registry (US)	36 MZ 18 DZ	33						
Wade *et al.* (1981)[c]	Female (*n* = 46)	19-58 ($\bar{x}$= 39.2)	3-day diet diary	Toronto Twin Register (Canada)	13 MZ 10 DZ	11 *ns*	-[f]	20 *ns*	66	48 *ns*	70	67
Heller *et al.* (1988)[c]	Male + female (*n* = 400)	17-66 ($\bar{x}$ = 36)	4-day food diary	Australia	106 MZ 94 DZ	38 *ns*	24 *ns* SFA: 10 *ns* MUFA: 33 *ns* PUFA: 3 *ns*	8 *ns*	31 *ns* SC: 20 *ns* CC: 55			
de Castro *et al.* (1993)[d]	Male + female (*n* = 390)	$\bar{x}$ = 38.8	7-day diet diary	Minnesota Twin Registry, (US)	109 MZ 86 DZ	65 (65)	51 (47)	57 (58)	61 (64)			
Hasselbalch *et al.* (2008)	Male & female (*n* = 1212)	18-67 ($\bar{x}$ = 38)	247-item FFQ	Geminakar (Denmark)	600							
	Males					38				36[g]	28	36
	Females					32				53[g]	55[g]	49
Hur *et al.* (1998)[e]	Male + female (*n* = 335)	18-77 ($\bar{x}$ = 42.4)	67-item FFQ	MSTRA (Multi-national)	66 MZ 51 DZ	32 (40)[e]	35 SFA: 37 PUFA: 46	16 *ns*	25 SC: 24 *ns* CC: 18 *ns*			

FFQ, food frequency questionnaire; USC, University of Southern California; MSTRA, Minnesota Study of Twins Reared Apart; MZ, monozygotic; DZ, dizygotic; *ns*, not significant (confidence interval spans 0); SFA, saturated fat; MUFA, monounsaturated fat; PUFA, polyunsaturated fat; CHO, carbohydrates; SC, simple carbohydrates; CC, complex carbohydrates.

*Heritability estimates are derived from the additive genetic effect value (a^2) calculated using structural equation modelling techniques as outlined by (Neale et al., 1992) unless otherwise specified. Twin pairs are same-sex. Ages are in years, unless otherwise specified. Energy intakes were calculated in kilocalories unless otherwise specified.

[a]Values are age- and sex- adjusted. Energy intake calculated in kilojoules.

[b]Age and sex-adjusted total caloric intake/meal heritability estimated only.

[c]Heritability estimates were calculated via Holzingers' *H* [(DZ variance-MZ variance)/DZ variance] or Falconer [2(rMZ-rDZ)] equations.

[d]Heritability estimated using the LISREL method. Heritability per meal is provided in paratheses.
[e]Twins reared apart study. Weight-adjusted heritability in parentheses (kilocalories/kg). Values are age- and gender- adjusted.
[f]MZ within-pair exceeded between-pair mean square.
[g]Heritability estimate derived from non-additive genetic effects (d^2) due to correlation coefficient between MZ two times greater than DZ twins.

Table 2. Selection of twin studies estimating heritability of empirically derived food intake patterns (from Pallister et al., 2014).*

Author	Sex	Age	Dietary assessment method	Twin registry	Twin pair count	Dietary factor	*Description*	h^2
Faith *et al.* (2008)[a]	Male & Female (*n* = 792)	7	24-h recall (parents)	MacArthur Longitudinal Study of Twins (US)	222 MZ 182 DZ	Peanut butter and jelly intake	Frequent intakes of legumes (peanuts and peanut butter), and jam and jelly.	*M*: 79
						Breakfast cereal and milk intake	Frequent intakes of milk and breakfast cereal.	*ns*
						Bread and butter intake	Frequent intakes of breads, butter and margarine.	*M*: 18 *F*: 20
						Adjusted fruit intake	Frequent intakes of non-citrus fruit, and fruit juice, punch and soda (reverse coded).	*M*: 26
						Adjusted red meat and pork intake	Frequent intakes of beef, pork, lamb and poultry (reverse coded).	*M*: 57
						Vegetable intake	Frequent intakes of deep-coloured vegetables and other vegetables.	*ns*
						Adjusted candy intake	Frequent intakes of candy, and sweets (reverse coded).	*M*: 41 *F*: 27
						Fish and lemon intake	Frequent intakes of fish and citrus fruit.	*M*: 12 *F*: 56
						High-salt snack food intake	Frequent intakes of high-salt snack foods.	*M*: 24
Keskitalo *et al.* (2007)[b]	Male + female (*n* = 4018)	22-27 ($\bar{x}$ = 24.4)	24-item FFQ	Finnish Twin Registry (Finland)	704 MZ 1490 DZ	"Healthy foods"	Frequent intakes of fresh vegetables, fruits, cooked vegetables, berries, porridge, muesli, cereals, reduced-fat cheeses, rice or pasta, chicken yoghurt, and fish.	*M*: 49 *F*: 54
						High-fat foods	Frequent intakes of fried foods, hamburgers, pizza, fried potatoes or French fries, creamy foods, and salty snacks.	*M*: 44 *F*: 47
						Sweet foods	Frequent intakes of other sweets, chocolate, and sweet desserts.	*M*: 42 *F*: 43
						Meat	High intakes of sausage and meat.	*M*: 39 *F*: 44
Teucher *et al.* (2007)[a]	Female (*n* = 3262)	18-79 ($\bar{x}$ = 48.1)	131-item FFQ	Twins UK (UK)	498 MZ 1133 DZ	Fruit and vegetable	Frequent intakes of fruit, allium and cruciferous vegetables; low intakes of fried potatoes.	43
						High alcohol	Frequent intakes of beer, wine and allium vegetables; low intakes of high fibre breakfast cereals and fruit.	48
						Traditional English	Frequent intakes of fried fish and potatoes, meats, savoury pies and cruciferous vegetables.	41
						Dieting	Frequent intakes of low-fat dairy products, low-sugar soda; low intakes of butter and sweet baked products.	41
						Low meat	Frequent intakes of baked beans, pizza and soy foods; low intakes of meat, other fish and seafood, and poultry.	43
Breen *et al.* (2006)[b,c]	Male + female (*n* = 428)	4-5 ($\bar{x}$ = 4.4)	95-item modified FFQ (parents)	Twins Early Development Study (UK)	103 MZ 111 DZ	Vegetables	High liking of broccoli, cabbage, carrots, cauliflower, green beans, mushrooms, onions, parsnips, salad greens and tomato.	37
						Desserts	High liking of cream, cakes, pastries, fruit pie, sponge pudding, custard and dairy desserts.	20

						Meat and fish	High liking of beef, lamb, pork, chicken, bacon, fried fish, white fish and oily fish.	78
						Fruit	High liking of apples, bananas, citrus fruits, grapes, peaches, strawberries and fruit juice.	51
Gunderson *et al.* (2006)[b,d]	Female (*n* = 700)	30-90 ($\bar{x}$ = 50)	100-item FFQ	Kaiser Permanente Twin Registry (US)	704 MZ 1490 DZ	'Healthy' dietary pattern	Frequent intakes of fish or chicken, carrots, tomatoes, salad, green or yellow vegetables, fruits, high fibre grains, rice and potatoes.	50
						'Unhealthy' dietary pattern	Frequent intakes of beef, pork, hot dogs, eggs, cheese, ice cream, butter, margarine, soda, and desserts.	*ns*
van den Bree *et al.* (1999)[b]	Male + female (*n* = 4640)	≥50	99-item semi-quantitative FFQ	Virginia Commonwealth University (US)	935 MZ 713 DZ	Less healthful	Frequent intakes of foods high in fat, salt and sugar.	33
						Healthful	Frequent intakes of a variety of vegetables, fruit, rice, yogurt, skim milk and dark bread.	33

FFQ, food frequency questionnaire; MZ, monozygotic; DZ, dizygotic; M, male; *ns*, not significant (confidence interval spans 0); F, female.
*Heritability estimates are derived from the additive genetic effect value (a^2) calculated using structural equation modelling techniques as outlined by (Neale et al., 1992) unless otherwise specified. Twin pairs are same-sex. Ages are in years, unless otherwise specified.
[a]Food patterns identified through principal component analysis.
[b]Food patterns identified through factor analysis.
[c]Heritabilities derived from a modified 95-item food frequency questionnaire which asked food 'liking' as opposed to intake frequency.
[d]Utilized a twins of mistaken zygosity approach.

Table 3. Selection of twin studies estimating heritability of food group intakes (from Pallister et al., 2014).*

Author	Sex	Age	Method	Twin registry	Twin pair count	Heritability													
						Veg	Potatoes	Fruit	Meat	Fish	Dairy	Eggs	Cereals	Legumes /nuts	Total	Sweets	Savoury	Fast food	Fats and oils
Pimpin *et al.* (2013)[a]	Male + female (*n* = 3605)	21 months	3-day diet diary	Gemini cohort (UK)	384 MZ 832 DZ	15	9	10	9 *ns*[h]		17	6 *ns*	9			5-15 *ns*[i]	4 *ns*		5 *ns*
Fildes *et al.* (2014)[b]	Male + Female (*n* = 2686)	3	114-item modified FFQ (parent)	Gemini cohort (UK)	458 MZ 872 DZ	54		53	48[h]		27[j]		32		29				
Hasselbalch *et al.* (2008)[c]	Male & female (*n* = 1212)	18-67 ($\bar{x}$ = 38)	247-item FFQ	Geminakar (Denmark)	600														
Males						24	68	-	34-47[e]	17	HF: 37 LF: 39	-	RG: 19 WG: 24			22 *ns*-45[e]			35-48[e]
Females						14 *ns*	28	-	29-38[i]	61	HF: 32 LF: 39	0 *ns*	RG: 12 *ns* WG: 20			23-27 *ns*[e]			9 *ns*-42[i]
Keskitalo *et al.* (2007)	Male + female (*n* = 4018)	22-27 ($\bar{x}$ = 24.4)	24-item FFQ	Finnish Twin Registry (Finland)	704 MZ 1490 DZ														
Males						38-40[e]	40-46[e]	37-51[e]	22-47[i]	45	38-48[i]	30	42-49[e]			23-55[i]	43	22-55[i]	
Females						48-50[e]	38-44[e]	44-39[e]	44-49[i]	44	37-43[i]	37	40-41[e]			33-54[i]	41	43-54[i]	
Teucher *et al.* (2007)[d]	Female (*n* = 3262)	18-79 ($\bar{x}$ = 48.1)	131-item FFQ	Twins UK (UK)	498 MZ 1133 DZ	35-46[f]	36[g]	40	29-39[e]	17 *ns*	HF: 24 LF: 36	29	RG: 8 WG: 29	30-32[e]		30	7 *ns*	26	

FFQ, food frequency questionnaire; MZ, monozygotic; DZ, dizygotic; *ns*, not significant (confidence interval spans 0); HF, high fat; LF, low fat; RG, refined grain; WG, whole grain.

*Heritability estimates are derived from the additive genetic effect value (a^2) calculated using structural equation modelling techniques as outlined by (Neale et al., 1992) unless otherwise specified. Twin pairs are same-sex. '*ns'* indicates confidence interval spans 0. '-' indicates model including additive genetic effects not the best fit. Ages are in years, unless otherwise specified. Values presented were derived from single variables, unless otherwise specified.

[a]Heritability analysis performed on age- and sex- adjusted residual scores.

[b]Heritability analysis performed on age- and sex- adjusted residual scores.

[c]Heritability analysis performed on food groups adjusted for total energy intake.

[d]Heritability analysis performed on energy-adjusted residual scores.

[e]Two variables are included in the range.

[f]Five variables are included in the range.

[g]Hertiability for root vegetable intake.

[h]Includes fish.

[i]Three variables are included in the range.

[j]Includes eggs.

Table 4. Selection of twin studies estimating heritability of fluid intakes (from Pallister et al., 2014).*

						Heritability							
Author	**Sex**	**Age**	**Dietary assessment method**	**Twin registry**	**Twin pair count**	**Water**	**Alcohol**	**Soda**	**Diet soda**	**Milk**	**Coffee**	**Tea**	**Fruit juice**
Pimpin *et al.* (2013)[a]	Male + female (*n* = 3605)	21 months	3-day diet diary	Gemini cohort (UK)	384 MZ 832 DZ	7		5 *ns*[h]		8			1 *ns*
de Castro *et al.* 1993[b]	Male + female (*n* = 390)	$\bar{x}$ = 38.8	7-day diet diary	Minnesota Twin Registry, (US)	109 MZ 86 DZ	80	73 *ns*	61	64	58 *ns*	73		-
Males						76[f]	82	38[i]	68[i]	-	71		-
Females						-	51	73	65[i]	52	73		-
Hasselbalch *et al.* (2008)[c]	Male & female (*n* = 1212)	18-67 ($\bar{x}$ = 38)	247-item FFQ	Geminakar (Denmark)	600								
Males								26			63[j]	63[j]	36
Females								30			-	-	0 *ns*
Teucher *et al.* (2007)[d]	Female (*n* = 3262)	18-79 ($\bar{x}$ = 48.1)	131-item FFQ	Twins UK (UK)	498 MZ 1133 DZ		28				41	38	0 *ns*
Hur *et al.* (1998)[e]	Male + female (*n* = 335)	18-77 ($\bar{x}$ = 42.4)	67-item FFQ	MSTRA (Multi-national)	66 MZ 51 DZ		38[g]	20 *ns*		33	29	46	25 *ns*

FFQ, food frequency questionnaire; MSTRA, Minnesota Study of Twins Reared Apart; MZ, monozygotic; DZ, dizygotic; *ns*, not significant (confidence interval spans 0).

*Heritability estimates are derived from the additive genetic effect value (a^2) calculated using structural equation modelling techniques as outlined by (Neale et al., 1992) unless otherwise specified. Twin pairs are same-sex. '-' indicates model including additive genetic effects not the best fit. Ages are in years, unless otherwise specified.

[a]Heritability analysis performed on age- and sex- adjusted residual scores.

[b]Heritability estimated using the LISREL method.

[c]Heritability analysis performed on food groups adjusted for total energy intake.

[d]Heritability analysis performed on energy-adjusted residual scores.

[e]Study used twins of mistaken zygosity approach. Heritability analysis performed on age- and gender- adjusted values.

[f]Significant dominant genetic effect

[g]Heritability of alcohol type intake.

[h]Heritability of other beverages group.

[i]Model including additive genetic effects not best fitting model.

[j]Obtained from group 'other', includes coffee, tea and unspecified items.

Appendix B. Chapter 3 Appendices

Document 1. Food frequency questionnaire before 2007

CONFIDENTIAL

RHEUMATOLOGY DEPARTMENT
OSTEOPOROSIS AND OSTEOARTHRITIS

TWIN RESEARCH

FOOD FREQUENCY QUESTIONNAIRE

This questionnaire asks for some background information about you especially what you eat.

Please answer every question. If you are uncertain about how to answer a question then do the best you can, but please do not leave a question blank.

Your answers will be treated as strictly confidential and will only be used for medical research.

PLEASE COMPLETE USING BLACK INK/PEN

STUDY NO.

DATE COMPLETED

YOUR DIET LAST YEAR

For each food there is an amount shown, either a 'medium serving' or a common household unit such as a slice or teaspoon. Please put a tick (✓) in the box to indicate how often, **on average**, you have eaten the specified amount of each food **during the past year.**

EXAMPLES

For white bread the amount is one slice, so if you ate 4 or 5 slices a day, you should put a tick in the column headed '4-5' per day.

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
Bread & savoury biscuits (one slice or biscuit)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
White bread and rolls								✓	

For chips, the amount is a 'medium serving', so if you had a helping of chips twice a week you should put a tick in the column headed '2-4 per week'.

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
Potatoes, Rice and Pasta (medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Chips				✓					

For very seasonal fruits such as strawberries and raspberries you should estimate your average use when the fruits are in season, so if you ate strawberries or raspberries about once a week when they are in season you should put a tick in the column headed 'once a week'.

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
Fruit (1 fruit medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Strawberries, raspberries kiwi fruit			✓						

Please estimate your average food use as best you can, and please answer every question - do not leave ANY lines blank. PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
MEAT AND FISH (medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Beef: roast, steak, mince, stew or casserole									
Beefburgers									
Pork: roast, chops, stew or slices									
Lamb: roast, chops or stew									
Chicken or other poultry eg. turkey									
Bacon									
Ham									
Corned beef, Spam, luncheon meats									
Sausages									
Savoury pies, eg. meat pie, pork pie, pasties, steak & kidney pie, sausage rolls									
Liver, liver paté, liver sausage									
Fried fish in batter, as in fish and chips									
Fish fingers, fish cakes									
Other white fish, fresh or frozen, eg. cod, haddock, plaice, sole, halibut									
Oily fish, fresh or canned, eg. mackerel, kippers, tuna, salmon, sardines, herring									
Shellfish, eg. crab, prawns, mussels									
Fish roe, taramasalata									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
BREAD AND SAVOURY BISCUITS (one slice or biscuit)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
White bread and rolls									
Brown bread and rolls									
Wholemeal bread and rolls									
Cream crackers, cheese biscuits									
Crispbread, eg. Ryvita									
CEREALS (one bowl)									
Porridge, Readybrek									
Breakfast cereal such as cornflakes, muesli etc.									
POTATOES, RICE AND PASTA (medium serving)									
Boiled, mashed, instant or jacket potatoes									
Chips									
Roast potatoes									
Potato salad									
White rice									
Brown rice									
White or green pasta, eg. spaghetti, macaroni, noodles									
Wholemeal pasta									
Lasagne, moussaka									
Pizza									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
DAIRY PRODUCTS AND FATS	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Single or sour cream (tablespoon)									
Double or clotted cream (tablespoon)									
Low fat yogurt, fromage frais (125g carton)									
Full fat or Greek yogurt (125g carton)									
Dairy desserts (125g carton)									
Cheese, eg. Cheddar, Brie, Edam (medium serving)									
Cottage cheese, low fat soft cheese (medium serving)									
Eggs as boiled, fried, scrambled, etc. (one)									
Quiche (medium serving)									
Low calorie, low fat salad cream(tablespoon)									
Salad cream, mayonnaise (tablespoon)									
French dressing (tablespoon)									
Other salad dressing (tablespoon)									
The following on bread or vegetables									
Butter (teaspoon)									
Block margarine, eg. Stork, Krona (teaspoon)									
Polyunsaturated margarine (tub), eg. Flora, sunflower (teaspoon)									
Other soft margarine,dairy spreads (tub), eg. Blue Band, Clover (teaspoon)									
Low fat spread (tub), eg. Outline, Gold (teaspoon)									
Very low fat spread (tub) (teaspoon)									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
SWEETS AND SNACKS (medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Sweet biscuits, chocolate , eg. digestive (one)									
Sweet biscuits, plain, eg. Nice, ginger (one)									
Cakes eg. fruit, sponge, home baked									
Cakes eg. fruit, sponge, ready made									
Buns, pastries eg. scones, flapjacks, home baked									
Buns, pastries eg. croissants, doughnuts, ready made									
Fruit pies, tarts, crumbles, home baked									
Fruit pies, tarts, crumbles, ready made									
Sponge puddings, home baked									
Sponge puddings, ready made									
Milk puddings, eg. rice, custard, trifle									
Ice cream, choc ices									
Chocolates, single or squares									
Chocolate snack bars eg. Mars, Crunchie									
Sweets, toffees, mints									
Sugar added to tea, coffee, cereal (teaspoon)									
Crisps or other packet snacks, eg. Wotsits									
Peanuts or other nuts									
SOUPS, SAUCES, AND SPREADS									
Vegetable soups (bowl)									
Meat soups (bowl)									
Sauces, eg. white sauce, cheese sauce, gravy (tablespoon)									
Tomato ketchup (tablespoon)									
Pickles, chutney (tablespoon)									
Marmite, Bovril (teaspoon)									
Jam, marmalade, honey (teaspoon)									
Peanut butter (teaspoon)									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
DRINKS	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Tea (cup)									
Coffee, instant or ground (cup)									
Coffee, decaffeinated (cup)									
Coffee whitener, eg. Coffee-mate (teaspoon)									
Cocoa, hot chocolate (cup)									
Horlicks, Ovaltine (cup)									
Wine (glass)									
Beer, lager or cider (half pint)									
Port, sherry, vermouth, liqueurs (glass)									
Spirits, eg. gin, brandy, whisky, vodka (single)									
Low calorie or diet fizzy soft drinks (glass)									
Fizzy soft drinks, eg. Coca cola, lemonade (glass)									
Pure fruit juice (100%) eg. orange, apple juice (glass)									
Fruit squash or cordial (glass)									
FRUIT (1 fruit or medium serving) **For very seasonal fruits such as strawberries, please estimate your average use when the fruit is in season**									
Apples									
Pears									
Oranges, satsumas, mandarins									
Grapefruit									
Bananas									
Grapes									
Melon									
Peaches, plums, apricots									
Strawberries, raspberries, kiwi fruit									
Tinned fruit									
Dried fruit, eg. raisins, prunes									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
VEGETABLES Fresh, frozen or tinned (medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Carrots									
Spinach									
Broccoli, spring greens, kale									
Brussels sprouts									
Cabbage									
Peas									
Green beans, broad beans, runner beans									
Marrow, courgettes									
Cauliflower									
Parsnips, turnips, swedes									
Leeks									
Onions									
Garlic									
Mushrooms									
Sweet peppers									
Beansprouts									
Green salad, lettuce, cucumber, celery									
Watercress									
Tomatoes									
Sweetcorn									
Beetroot									
Coleslaw									
Avocado									
Baked beans									
Dried lentils, beans, peas									
Tofu , soya meat, TVP, Vegeburger									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

YOUR DIET LAST YEAR, continued

2. Are there any **OTHER** foods which you ate more than once a week? Yes ☐ No ☐

If yes, please list below

Food	Usual serving size	Number of times eaten each week

3. What type of milk did you most often use?

Select one only

Full cream, silver ☐
Semi-skimmed, red/white ☐
Skimmed/blue ☐
Channel Islands, gold ☐
Dried milk ☐
Soya ☐
Other, *specify* ☐
None ☐

4. How much milk did you drink each day, including milk with tea, coffee, cereals etc?

None ☐
Three quarters of a pint ☐
Quarter of a pint ☐
One pint ☐
Half a pint ☐
More than one pint ☐

5. Did you usually eat breakfast cereal (excluding porridge and Ready Brek mentioned earlier)? Yes ☐ No ☐

If yes, which brand and type of breakfast cereal, including muesli, did you usually eat?

List the one or two types most often used

Brand	Type

6. What kind of fat did you most often use for frying, roasting, grilling etc?

Select one only

Butter ☐
Solid vegetable fat ☐
Lard/dripping ☐
Margarine ☐
Vegetable oil ☐
None ☐

If you used vegetable oil, please give type eg. corn, sunflower ☐

7. What kind of fat did you most often use for baking cakes etc?

Select one only

Butter ☐
Solid vegetable fat ☐
Lard/dripping ☐
Margarine ☐
Vegetable oil ☐
None ☐

If you used margarine, please give name or type eg. Flora, Stork ☐

8. How often did you eat food that was fried at home?

Daily ☐ 1-3 times a week ☐ 4-6 times a week ☐
Less than once a week ☐ Never ☐

9. How often did you eat fried food away from home?

Daily ☐ 1-3 times a week ☐ 4-6 times a week ☐
Less than once a week ☐ Never ☐

10. What did you do with the visible fat on your meat?

Ate most of the fat ☐ Ate as little as possible ☐
Ate some of the fat ☐ Did not eat meat ☐

11. How often did you eat grilled or roast meat? ☐☐ times a week

12. How well cooked did you usually have grilled or roast meat?

Well done /dark brown ☐ Lightly cooked/rare ☐
Medium ☐ Did not eat meat ☐

13. How often did you add salt to food while cooking?

Always ☐ Rarely ☐
Usually ☐ Never ☐
Sometimes ☐

14. How often did you add salt to any food at the table?

Always ☐ Rarely ☐
Usually ☐ Never ☐
Sometimes ☐

15. Did you regularly use a salt substitute (eg LoSalt)? Yes ☐ No ☐

If yes, which brand? ☐

Have you taken any vitamins, minerals, fish oils, fibre or other food supplements during the past year?

YES ☐ NO ☐ DON'T KNOW ☐

If yes, please complete the table below. If you have taken more than 5 types of supplement please put the most frequently consumed brands first.

Vitamin supplements	**Average frequency** Tick **one** box per line to show how often on average you consumed supplements									
Name and brand Please list full name, **brand** and strength	Dose Please state number of pills, capsules or teaspoons consumed	Never or less than once a month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Thank you for your help

Document 2: Food frequency questionnaire after 2007

Study Number: | | | | | | |

Twin Research & Genetic Epidemiology Unit

King's College London, St Thomas' Campus,
Lambeth Palace Road, London SE1 7EH
Tel: 0207 188 5555, Fax: 0207 188 6761
Email: twinrecruitment@gstt.nhs.uk
Website: www.twinsuk.ac.uk

June 2007

Dear Twin,

We are excited to send you a new questionnaire about food and nutrition developed with our colleagues at the University of East Anglia in Norwich.

The aim of this important study is to identify the genetic variations that influence our individual response to the food we eat, including vitamins and minerals. We are also including questions about your general health, which may be affected by your diet.

You may have answered some of the questions several years ago, but we need to keep our information as up to date as possible, so please do complete all the questions asked. We would like to build up a full profile of how diet may be related to health status. Please complete the questionnaire using a black pen. If you have any queries regarding this questionnaire please contact Dr Paula Skidmore on 01603 591270.

Thank you very much for your time and your continued support. Please return your completed questionnaire in the pre-paid envelope enclosed.

Yours sincerely,

Professor Tim Spector
Director, Twin Research & Genetic Epidemiology Unit

Study Number: | | | | | | |

YOUR DIET LAST YEAR

For each food there is an amount shown, either a "medium serving" or an example of a serving such as "a slice" or "a teaspoon". Please put a cross (×) in the box to indicate how often, **on average**, you have eaten a specified amount of food over the last 12 months. If you complete a box incorrectly please fill the wrong one. e.g. ■ and put a cross in the correct box.

The rest of this page shows examples of how to fill in the questionnaire.

These questions are all about the food you have eaten over the last year.

EXAMPLE 1

For white bread the serving size is one slice or roll, so if you ate 2 to 4 slices a week on average, over the last 12 months, you should put a cross in the column headed "2-4 per week".

FOODS AND AMOUNTS	AVERAGE USE IN THE LAST TWELVE MONTHS								
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
BREAD (one slice or roll)									
White bread and rolls	☐	☐	☐	☒	☐	☐	☐	☐	■ example of mistake

EXAMPLE 2

For chips, the serving size is a small side plate so if you ate a serving of chips twice a week, over the last 12 months, you should put a cross in the column headed "2-4 per week".

FOODS AND AMOUNTS	AVERAGE USE IN THE LAST TWELVE MONTHS								
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
POTATOES (one small side plate)									
Chips	☐	☐	☐	☒	☐	☐	☐	☐	☐

EXAMPLE 3

For very seasonal fruits such as strawberries and raspberries you should estimate your average intake in summer, so if you ate strawberries or raspberries about once a week in summer you should put a cross in the column headed "once a week".

FOODS AND AMOUNTS	AVERAGE USE IN THE LAST TWELVE MONTHS								
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
FRUIT (1 fruit or handful)									
Strawberries, raspberries, kiwi fruit	☐	☐	☒	☐	☐	☐	☐	☐	☐

Study Number: ☐☐☐☐☐☐

START OF QUESTIONNAIRE

1a. On average do you think you eat the same, more or less than someone of your own age and sex? **Please put a cross (×) in only one box.** ☐ Same ☐ More ☐ Less

1b. On an average day how many portions of fruit do you eat? ☐

1c. On an average day how many portions of vegetables do you eat? ☐

Please estimate your average food use as best you can, and please answer every question - do not leave ANY lines blank. **PLEASE PUT A CROSS (×) ON EVERY LINE**

FOODS AND AMOUNTS	AVERAGE USE IN THE LAST TWELVE MONTHS								
DAIRY PRODUCTS AND FATS	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Single or sour cream (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Double or clotted cream (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Low fat yoghurt, fromage frais (small pot)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Full fat or Greek yoghurt (small pot)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Dairy desserts (small pot) e.g. chocolate mousse, cream caramels	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cheese, e.g. cheddar, brie, edam (matchbox size)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Low fat cheese e.g. reduced fat cheddar (matchbox size)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cottage cheese, low fat soft cheese (2 tablespoons)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Eggs as boiled, fried, scrambled, etc. (one)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Quiche (slice)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Low calorie, low fat salad cream (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Full fat salad cream, mayonnaise (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
French dressing (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Other salad dressing (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐

Please check that you have a cross (×) on EVERY line

Study Number: ☐☐☐☐☐☐

FOODS AND AMOUNTS	AVERAGE USE IN THE LAST TWELVE MONTHS								
The following list refers to **DAIRY PRODUCTS AND FATS** that you put on bread or cooked vegetables	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Butter (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Reduced fat butter (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Block margarine, e.g. Stork, Krona (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Polyunsaturated margarine, e.g. Flora, sunflower (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Olive oil spread (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Other soft margarine, dairy spreads, e.g. Blue Band, Clover (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Low fat spread, e.g. Outline, Gold (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Very low fat spread (teaspoon) e.g. Diet Flora	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cholesterol lowering fat spreads e.g. Benecol (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐

FRUIT - For seasonal fruits marked*, please estimate your average use in summer

Apples (1 fruit)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Pears (1 fruit)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Oranges, satsumas, mandarins (1 fruit)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Grapefruit (half)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Bananas (1fruit)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Grapes (handful)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Melon (1 slice)	☐	☐	☐	☐	☐	☐	☐	☐	☐
* Peaches, plums, apricots (1 fruit)	☐	☐	☐	☐	☐	☐	☐	☐	☐
* Strawberries, raspberries, other berries, kiwi fruit (one fruit or handful)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Tinned fruit (handful)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Dried fruit, e.g. raisins, prunes (heaped tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐

Please check that you have a cross (☒) on EVERY line

Study Number: | | | | | | |

FOODS AND AMOUNTS | **AVERAGE USE IN THE LAST TWELVE MONTHS**

VEGETABLES Fresh, frozen or tinned (handful)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Carrots	☐	☐	☐	☐	☐	☐	☐	☐	☐
Spinach	☐	☐	☐	☐	☐	☐	☐	☐	☐
Broccoli, spring green, kale	☐	☐	☐	☐	☐	☐	☐	☐	☐
Brussel sprouts	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cabbage	☐	☐	☐	☐	☐	☐	☐	☐	☐
Peas	☐	☐	☐	☐	☐	☐	☐	☐	☐
Green beans, broad beans, runner beans	☐	☐	☐	☐	☐	☐	☐	☐	☐
Baked beans	☐	☐	☐	☐	☐	☐	☐	☐	☐
Marrow, courgettes	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cauliflower	☐	☐	☐	☐	☐	☐	☐	☐	☐
Parsnips, turnips, swedes	☐	☐	☐	☐	☐	☐	☐	☐	☐
Leeks	☐	☐	☐	☐	☐	☐	☐	☐	☐
Onions	☐	☐	☐	☐	☐	☐	☐	☐	☐
Garlic (clove)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Mushrooms	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sweet peppers	☐	☐	☐	☐	☐	☐	☐	☐	☐
Beansprouts	☐	☐	☐	☐	☐	☐	☐	☐	☐
Green salad, lettuce, cucumber, celery	☐	☐	☐	☐	☐	☐	☐	☐	☐
Watercress	☐	☐	☐	☐	☐	☐	☐	☐	☐
Tomatoes	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sweetcorn	☐	☐	☐	☐	☐	☐	☐	☐	☐
Beetroot	☐	☐	☐	☐	☐	☐	☐	☐	☐
Coleslaw	☐	☐	☐	☐	☐	☐	☐	☐	☐
Avocado	☐	☐	☐	☐	☐	☐	☐	☐	☐
Pulses e.g. lentils, beans, peas	☐	☐	☐	☐	☐	☐	☐	☐	☐
Meat substitutes e.g. tofu, soyameat, textured vegetable protein, vegeburger	☐	☐	☐	☐	☐	☐	☐	☐	☐

Please check that you have a cross (×) on EVERY line

Study Number: | | | | | | |

FOODS AND AMOUNTS **AVERAGE USE IN THE LAST TWELVE MONTHS**

MEAT AND FISH (half a small side plate)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Beef: roast, steak, mince, stew or casserole	☐	☐	☐	☐	☐	☐	☐	☐	☐
Beefburgers	☐	☐	☐	☐	☐	☐	☐	☐	☐
Pork: roast, chops or stew	☐	☐	☐	☐	☐	☐	☐	☐	☐
Lamb: roast, chops or stew	☐	☐	☐	☐	☐	☐	☐	☐	☐
Chicken or other poultry e.g. turkey	☐	☐	☐	☐	☐	☐	☐	☐	☐
Bacon or gammon	☐	☐	☐	☐	☐	☐	☐	☐	☐
Ham, cured meats & chorizo	☐	☐	☐	☐	☐	☐	☐	☐	☐
Corned Beef, Spam, luncheon meats	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sausages	☐	☐	☐	☐	☐	☐	☐	☐	☐
Savoury pies, e.g. meat pie, pork pie, pasties, steak & kidney pie, sausage rolls	☐	☐	☐	☐	☐	☐	☐	☐	☐
Liver, liver pate, liver sausage	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fried fish in batter, as in fish and chips	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fish fingers, fish cakes & breaded fish	☐	☐	☐	☐	☐	☐	☐	☐	☐
Other white fish, fresh or frozen, e.g. cod, plaice, sole, haddock, halibut	☐	☐	☐	☐	☐	☐	☐	☐	☐
Oily fish, fresh or canned, e.g. tuna, mackerel, kippers, salmon, sardines, herring	☐	☐	☐	☐	☐	☐	☐	☐	☐
Shellfish, e.g. crab, prawns, mussels	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fish roe, taramasalata	☐	☐	☐	☐	☐	☐	☐	☐	☐

Please check that you have a cross (×) on EVERY line

Study Number: | | | | | | |

FOODS AND AMOUNTS	AVERAGE USE IN THE LAST TWELVE MONTHS								
BREAD AND SAVOURY BISCUITS (one slice, or roll, or biscuit)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
White bread/rolls	☐	☐	☐	☐	☐	☐	☐	☐	☐
Brown bread/rolls	☐	☐	☐	☐	☐	☐	☐	☐	☐
Wholemeal & granary bread/rolls	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cream crackers, savoury biscuits	☐	☐	☐	☐	☐	☐	☐	☐	☐
Crispbread, e.g. Ryvita	☐	☐	☐	☐	☐	☐	☐	☐	☐
Naan, poppadoms, flour tortillas	☐	☐	☐	☐	☐	☐	☐	☐	☐
SOUPS, SAUCES AND SPREADS									
Vegetable soups (bowl)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Meat soups (bowl) (to include meat and vegetable soups)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sauces, e.g. white sauce, cheese sauce, gravy (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Tomato ketchup (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Pickles, chutney (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Marmite, Bovril (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Jam, marmalade, honey (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Peanut butter (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐

Please check that you have a cross (☒) on EVERY line

Study Number: | | | | | | |

FOODS AND AMOUNTS **AVERAGE USE IN THE LAST TWELVE MONTHS**

SWEETS AND SNACKS (medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Cakes e.g. fruit, sponge, home baked	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cakes e.g. fruit, sponge, ready made	☐	☐	☐	☐	☐	☐	☐	☐	☐
Buns, pastries e.g. scones, flapjacks, croissants, doughnuts, home baked	☐	☐	☐	☐	☐	☐	☐	☐	☐
Buns, pastries e.g. scones, flapjacks, croissants, doughnuts, ready made	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fruit pies, tarts, crumbles, home baked	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fruit pies, tarts, crumbles, ready made	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sponge puddings, home baked	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sponge puddings, ready made	☐	☐	☐	☐	☐	☐	☐	☐	☐
Milk puddings e.g. rice, custard, trifle	☐	☐	☐	☐	☐	☐	☐	☐	☐
Ice cream, choc ices	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sweet biscuits, chocolate, e.g. digestive (one)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sweet biscuits, plain, e.g. Nice, ginger (one)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Reduced fat biscuits e.g. Go Ahead, Highlights (one small packet or one small bar/biscuit)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cereal bars (one)	☐	☐	☐	☐	☐	☐	☐	☐	☐
White or milk chocolates, single or squares (one)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Dark chocolates, single or squares (one)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Chocolate snack bars e.g. Mars, Crunchie (one)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sweets, toffees, mints (small packet)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sugar added to tea, coffee, cereal (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐

Please check that you have a cross (✕) on EVERY line

Study Number: | | | | | | |

FOODS AND AMOUNTS	AVERAGE USE IN THE LAST TWELVE MONTHS								
SWEETS AND SNACKS	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Crisps or other packet snacks, e.g. Wotsits (one packet)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Salted nuts e.g. peanuts, cashews (handful)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Unsalted nuts, e.g. brazil, walnuts (handful)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Seeds e.g. Sunflower, pumpkin (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
CEREAL (one bowl)									
Porridge, Readybreak, oats	☐	☐	☐	☐	☐	☐	☐	☐	☐
Breakfast cereal e.g. Cornflakes, Rice Krispies	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sugar topped cereals e.g. Frosties	☐	☐	☐	☐	☐	☐	☐	☐	☐
Muesli	☐	☐	☐	☐	☐	☐	☐	☐	☐
High Fibre cereals e.g. Branflakes, All Bran, Fruit and Fibre	☐	☐	☐	☐	☐	☐	☐	☐	☐
POTATOES, RICE AND PASTA (one small side plate)									
Boiled, mashed, instant or one jacket potato	☐	☐	☐	☐	☐	☐	☐	☐	☐
Chips, roast potatoes	☐	☐	☐	☐	☐	☐	☐	☐	☐
Potato salad	☐	☐	☐	☐	☐	☐	☐	☐	☐
White rice	☐	☐	☐	☐	☐	☐	☐	☐	☐
Brown rice	☐	☐	☐	☐	☐	☐	☐	☐	☐
White or green pasta, e.g. spaghetti, macaroni, noodles	☐	☐	☐	☐	☐	☐	☐	☐	☐
Wholemeal pasta	☐	☐	☐	☐	☐	☐	☐	☐	☐
Lasagne, moussaka	☐	☐	☐	☐	☐	☐	☐	☐	☐
Pizza (one slice)	☐	☐	☐	☐	☐	☐	☐	☐	☐

Please check that you have a cross (×) on EVERY line

Study Number: | | | | | | |

FOODS AND AMOUNTS **AVERAGE USE IN THE LAST TWELVE MONTHS**

DRINKS	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Tea (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Green tea (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fruit tea (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Coffee, instant or ground (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Coffee, decaffeinated (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Coffee whitener, e.g. Coffee-mate (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cocoa, hot chocolate (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Low fat hot chocolate (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Horlicks, Ovaltine (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
White wine (small glass)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Red wine (small glass)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Beer, lager or cider (half pint)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Port, sherry, vermouth, liqueurs (pub measure)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Spirits, e.g. gin, brandy, whisky, vodka (pub measure)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Low calorie or diet fizzy soft drinks (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fizzy soft drinks, e.g. Coca Cola, lemonade (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Pure fruit juice (100%) e.g. orange, apple juice (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fruit squash or cordial (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Smoothies (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐

Please check that you have a cross (×) on EVERY line

Study Number: | | | | | | |

YOUR DIET LAST YEAR, continued

2. Are there any **OTHER** foods which you ate more than once a week over the last year?

No ☐ ⇨ **GO TO** question 3 Yes ☐ **If yes**, please list below

Food	Usual serving size	Number of times eaten each week

3. Have you drunk milk or a milk substitute over the past year? (including in hot drinks)

No ☐ ⇨ **GO TO** question 6 Yes ☐ ⇨ **GO TO** question 4

4. What type of milk or milk substitute did you most often use over the last year?

Select one only

Full cream	☐	Semi-skimmed	☐
Skimmed	☐	Channel Islands	☐
Dried milk	☐	Soya	☐

Other (please specify) _____

5. About how much milk did you drink each day, including milk with tea, coffee, cereals etc?

Less than a quarter of a pint	☐	One Pint	☐
Half a pint	☐	More than one pint	☐
Three quarters of a pint	☐		

6. Did you usually eat breakfast cereal in the last year (excluding Porridge and Ready Brek mentioned earlier)?

No ☐ ⇨ **GO TO** question 7 Yes ☐ **If yes,** which brand and type of breakfast cereal, including muesli, did you usually eat?

List the one or two types most often used

Brand e.g. Kellogg's	Type e.g. cornflakes

7. Do you use oil or fat for frying, roasting, grilling etc?

No ☐ ⇨ **GO TO** question 9 Yes ☐ ⇨ **GO TO** question 8

Study Number: ☐☐☐☐☐☐

8. What kind of fat did you most often use for frying, roasting, grilling etc? **Select one only**

Butter ☐ Olive oil ☐ Margarine ☐ Vegetable oil ☐
Lard/dripping ☐ Solid vegetable fat ☐ Low fat spray ☐ Other ☐

If you used vegetable oil or other oil
please give type e.g. corn, sunflower, sesame _____

9. What kind of fat did you most often use for baking cakes etc? **Select one only**

I don't bake ☐ Lard/dripping ☐ Solid vegetable fat ☐ None ☐
Butter ☐ Vegetable oil ☐ Margarine ☐

If you used margarine, please give name or type e.g. Flora, Stork _____

10. How often did you eat food that was fried at home?

Never ☐ Less than once a week ☐ 1-3 times a week ☐
4-6 times a week ☐ Daily ☐

11. How often did you eat fried food away from home?

Never ☐ Less than once a week ☐ 1-3 times a week ☐
4-6 times a week ☐ Daily ☐

12. Have you eaten meat in the last year?

No ☐ ⇨ **GO TO** question 17 Yes ☐ ⇨ **GO TO** question 13

13. What did you do with the visible fat on your meat?

Did not eat meat with visible fat ☐ Ate some of the fat ☐
Ate as little as possible ☐ Ate most of the fat ☐

14. Did you eat roast/grilled meat over the past year?

No ☐ ⇨ **GO TO** question 17 Yes ☐ ⇨ **GO TO** question 15

15. How often did you eat grilled or roast meat?

Daily ☐ 1-3 times a week ☐
4-6 times a week ☐ Less that once a week ☐

Study Number: ☐☐☐☐☐☐

16. How well cooked did you usually have grilled or roast meat?

Well done / dark brown ☐ Lightly cooked / rare ☐
Medium ☐

17. How often did you add salt to food while cooking?

Never ☐ Sometimes ☐ Always ☐
Rarely ☐ Usually ☐

18. How often did you add salt to any food at the table?

Never ☐ Sometimes ☐ Always ☐
Rarely ☐ Usually ☐

19. Do you regularly use a salt substitute (e.g. LoSalt)?

No ☐ Yes ☐ **If yes,** which brand? _____

20. We would like to summarise what you have told us. During the course of last year, on average, how many times a week did you eat the following foods?

Food Type	**Times/week**	**Portion size**
Vegetables (excluding potatoes but including baked beans)	☐☐	medium serving
Salads	☐☐	medium serving
Fruit and fruit products (not including fruit juice)	☐☐	medium serving or 1 fruit
Oily fish and fish products	☐☐	medium serving
Non oily fish and fish products	☐☐	medium serving
Meat, meat products and meat dishes (including bacon, ham and chicken)	☐☐	medium serving

21. Have you taken any vitamins, minerals, fish oils, fibre or other food supplements during the past year?

No ☐ ⇨ **GO TO** section on **HEALTH** on page 15 Yes ☐ ⇨ **GO TO** question 22

Study Number: | | | | | | |

22. Please list any vitamins, minerals, fish oils, fibre or other food supplements taken during the past year. If you have taken more than 6 types of supplement please list the 6 most frequently consumed brands first.

Vitamin supplements

Average frequency
Cross **one** box per line to show how often on average you consumed supplements

Name and brand Please list full name and **brand**	**Number** Please state number of pills, capsules or teaspoons consumed	Never or less than once a month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per say	4-5 per day	6+ per day
	1 - 9	☐	☐	☐	☐	☐	☐	☐	☐	☐
	1 - 9	☐	☐	☐	☐	☐	☐	☐	☐	☐
	1 - 9	☐	☐	☐	☐	☐	☐	☐	☐	☐
	1 - 9	☐	☐	☐	☐	☐	☐	☐	☐	☐
	1 - 9	☐	☐	☐	☐	☐	☐	☐	☐	☐
	1 - 9	☐	☐	☐	☐	☐	☐	☐	☐	☐

Appendix C. Chapter 4 Appendices

Table 1. Study population characteristics

	Whole population (*n*=3559)		Discordant monozygotic twins		
Variable	**Mean (SD)**	**Range**	***n* pairs**	**Mean**	**Range**
Age (years)	55.3 (13.4)	18 - 84			
Body mass index (kg/m^2)	26.1 (4.9)	13.6 - 50.1			
Intakes (servings/week)					
Allium vegetables	4.5 (4.2)	0 - 84.5	174	7.1 (5.6)	0 - 37
Apples/pears	4.7 (4.9)	0 - 74	206	7.0 (6.2)	0 - 36
Avocado	0.4 (0.9)	0 - 18	129	1.6 (2)	0 - 18
Baked beans	0.9 (1.0)	0 - 7	149	1.8 (1.6)	0 - 7
Bananas	3.9 (3.7)	0 - 42	210	5.3 (5)	0 - 32
Beef burgers	0.1 (0.3)	0 - 3	172	0.3 (0.4)	0 - 3
Beer	0.7 (2.4)	0 - 42	63	3.5 (5.6)	0 - 32
Berries	0.9 (1.4)	0 - 14	137	2.1 (2.7)	0 - 14
Black tea	18.9 (13.8)	0 - 42	240	18.4	0 - 42
Butter	3.5 (6.1)	0 - 50	173	8.5 (8.9)	0 - 42
Cheese	3.1 (3.1)	0 - 36	198	5.0 (4.8)	0 - 36
Chocolate	3.8 (6.1)	0 - 84	153	8.9 (10.7)	0 - 84
Citrus fruit	3.0 (4.3)	0 - 45	167	6.2 (6.6)	0 - 36
Coffee	12.3 (12.4)	0 - 84	252	17.4	0 - 84
Cooked potatoes	3.4 (2.6)	0 - 42	207	4.1 (3.3)	0 - 32
Cream	0.5 (1.3)	0 - 36	70	2.1 (2.1)	0 - 8
Crispbread	1.2 (3.4)	0 - 42	85	6.1 (8.6)	0 - 42
Cruciferous vegetables	6.1 (4.7)	0 - 98	219	8.6 (5.7)	0 - 35
Dairy desserts	0.2 (0.7)	0 - 18	94	0.9 (1.8)	0 - 18
Eggs	1.4 (1.6)	0 - 18	209	2.6 (2.5)	0 - 18
Fried fish	0.5 (0.7)	0 - 11	195	0.9 (1)	0 - 7.5
Fried potatoes	1.4 (1.4)	0 - 32	184	2.2 (1.7)	0 - 12
Fruit juice	3.4 (4.3)	0 - 42	221	5.8 (6)	0 - 42
Grapes	2.0 (2.9)	0 - 42	161	4.3 (4.7)	0 - 32
Green leafy vegetables	4.7 (4.4)	0 - 53	186	8.0 (7)	0 - 43
Herbal tea	1.6 (5.2)	0 - 64	117	9.5 (10.3)	0 - 50
High fat milk	1.0 (4.0)	0 - 45.5	59	7.4 (10.4)	0 - 45.5
High fibre breakfast	2.7 (3.5)	0 - 42	220	4.1 (4.1)	0 - 32.5
High sugar drinks	2.5 (5.6)	0 - 60	109	9.1 (10.2)	0 - 49
Ice cream	0.7 (1.2)	0 - 18	132	2.0 (1.8)	0 - 7
Lasagne	0.3 (0.4)	0 - 3	307	0.4 (0.4)	0 - 3
Legumes	4.3 (3.3)	0 - 43	230	5.8 (4.2)	0 - 32
Low fat milk	3.3 (2.4)	0 - 10.5	228	3.6 (3)	0 - 10.5
Low fibre breakfast cereals	1.6 (2.9)	0 - 42	224	3.4 (3.9)	0 - 42
Low fat spread	1.4 (4.5)	0 - 74	90	6.8 (8.5)	0 - 42
Malt drinks	0.3 (1.4)	0 - 32	66	2.9 (3.1)	0 - 18
Margarine	2.3 (4.9)	0 - 64	186	6.5 (7.8)	0 - 42.5
Marrow	0.7 (1.3)	0 - 32	130	2.1 (1.8)	0 - 7
Meat	2.4 (2.0)	0 - 16.5	222	3.3 (2.4)	0 - 11
Melon	0.9 (1.7)	0 - 32	150	2.8 (3.4)	0 - 32
Mushrooms	1.5 (1.7)	0 - 32	258	2.4 (1.8)	0 - 7
Nuts	1.7 (3.2)	0 - 43.5	144	5.0 (5.3)	0 - 41.5
Oily fish	1.1 (1.2)	0 - 18	217	2.1 (1.7)	0 - 18
Other fish/seafood	1.2 (1.2)	0 - 16.5	171	2.1 (1.7)	0 - 9.5
Peaches	0.7 (1.3)	0 - 14	141	1.8 (2.2)	0 - 14
Pizza	0.5 (0.7)	0 - 8	199	0.8 (0.9)	0 - 5.5
Polyunsaturated margarine	3.1 (6.0)	0 - 42	143	9.3 (10)	0 - 42
Porridge	1.4 (2.4)	0 - 32	215	3.0 (2.7)	0 - 7
Poultry	1.9 (1.3)	0 - 7	344	2.2 (1.4)	0 - 7
Processed meats	1.8 (3.1)	0 - 60	190	4.3 (4.2)	0 - 32
Processed fruit	2.8 (2.3)	0 - 18	228	3.9 (2.8)	0 - 15.5
Root vegetables	4.6 (3.7)	0 - 96	202	6.3 (4.5)	0 - 28
High fat salad dressing	1.6 (2.2)	0 - 25	201	3.1 (3.3)	0 - 25
Low fat salad dressing	0.7 (1.4)	0 - 18	140	2.3 (2.4)	0 - 18

Table 1. Study population characteristics

	Whole population (*n*=3559)		Discordant monozygotic twins		
Variable	**Mean (SD)**	**Range**	***n* pairs**	**Mean**	**Range**
Savoury pies	0.3 (0.5)	0 - 7	163	0.5 (0.8)	0 - 7
Savoury snacks	2.1 (3.0)	0 - 35	170	4.5 (5.3)	0 - 35
Seasonings	3.6 (3.7)	0 - 84.5	195	5.6 (6.6)	0 - 84.5
Low sugar soda	1.8 (4.8)	0 - 42	92	7.9 (9.2)	0 - 42
Soup	1.1 (1.9)	0 - 64	149	2.9 (2.6)	0 - 14
Soy/other milk	0.1 (0.8)	0 - 10.5	68	1.7 (1.9)	0 - 10.5
Soy foods	0.2 (1.0)	0 - 18	60	2.2 (3.2)	0 - 18
Spirits/liquor	1.2 (3.2)	0 - 50	97	5.1 (7)	0 - 42.5
Sweet baked products	7.9 (8.9)	0 - 85	155	15.5	0 - 76.5
Sweetcorn	0.9 (1.3)	0 - 18	187	2.1 (1.9)	0 - 18
Sweet peppers	1.4 (1.8)	0 - 32	245	2.5 (2.1)	0 - 18
Confectionary/jam	8.1 (11.9)	0 - 84.5	149	18.2	0 - 84
Tomatoes	4.0 (3.6)	0 - 42	175	6.5 (6.2)	0 - 42
White/brown bread, refined	8.3 (8.5)	0 - 70	206	12.8	0 - 43.5
Wholemeal bread/grains	4.9 (6.7)	0 - 44	179	10.3 (9.2)	0 - 43
Wine	3.9 (6.1)	0 - 64	111	12.2	0 - 43
Yoghurt	2.9 (3.7)	0 - 47.5	180	5.3 (5.6)	0 - 47.5

Table 2. Food frequency questionnaire variables included in each food group and descriptions

Food group	FFQ Variable Name	FFQ Description	Included FFQ1	Included FFQ2
Allium vegetables	Garlic	Garlic (clove)	TRUE	TRUE
Allium vegetables	Leeks	Leeks	TRUE	TRUE
Allium vegetables	Onions	Onions	TRUE	TRUE
Apples/pears	Apples	Apples (1 fruit)	TRUE	TRUE
Apples/pears	Pears	Pears (1 fruit)	TRUE	TRUE
Avocado	Avocado	Avocado	TRUE	TRUE
Baked beans	BakedBeans	Baked beans	TRUE	TRUE
Bananas	Bananas	Bananas (1fruit)	TRUE	TRUE
Beef burgers	MFburgers	Beefburgers	TRUE	TRUE
Beer	Beer	Beer, lager or cider (half pint)	TRUE	TRUE
Berries	Strawberries	Strawberries, raspberries, other berries, kiwi fruit (one fruit or handful)	TRUE	TRUE
Black tea	Tea	Tea (cup)	TRUE	TRUE
Butter	RedFatButter	Reduced fat butter (teaspoon)	FALSE	TRUE
Butter	Butter	Butter (teaspoon)	TRUE	TRUE
Cheese	DairyLFcheese	Low fat cheese e.g. reduced fat cheddar (matchbox size)	FALSE	TRUE
Cheese	Dairycheese	Cheese, e.g. cheddar, brie, edam (matchbox size)	TRUE	TRUE
Cheese	DairyCottageCheese	Cottage cheese, low fat soft cheese (2 tablespoons)	TRUE	TRUE
Chocolate	ChocsDark	Dark chocolates, single or squares (one)	FALSE	TRUE
Chocolate	ChocsMilk	White or milk chocolates, single or squares (one)	FALSE	TRUE
Chocolate	LowFatHotChoc	Low fat hot chocolate (cup)	FALSE	TRUE
Chocolate	Cocoa	Cocoa, hot chocolate (cup)	TRUE	TRUE
Chocolate	Chocolatebar	Chocolate snack bars e.g. Mars, Crunchie (one)	TRUE	TRUE
Chocolate	Chocolates	Chocolate, fancy and filled	TRUE	FALSE
Citrus fruit	Grapefruit	Grapefruit (half)	TRUE	TRUE
Citrus fruit	Oranges	Oranges, satsumas, mandarins (1 fruit)	TRUE	TRUE
Coffee	Coffee	Coffee, instant or ground (cup)	TRUE	TRUE
Coffee	CoffeeDecaffeinated	Coffee, decaffeinated (cup)	TRUE	TRUE
Confectionary/jam	Jam	Jam, marmalade, honey (teaspoon)	TRUE	TRUE
Confectionary/jam	SugarAdded	Sugar added to tea, coffee, cereal (teaspoon)	TRUE	TRUE
Confectionary/jam	SweetsToffees	Sweets, toffees, mints (small packet)	TRUE	TRUE
Cooked potatoes	Boiledpotato	Boiled, mashed, instant or one jacket potato	TRUE	TRUE
Cream	Dairydouble	Double or clotted cream (tablespoon)	TRUE	TRUE
Cream	Dairysingle	Single or sour cream (tablespoon)	TRUE	TRUE
Crispbread	Crispbread	Crispbread, e.g. Ryvita	TRUE	TRUE
Cruciferous vegetables	Broccoli	Broccoli, spring green, kale	TRUE	TRUE
Cruciferous vegetables	BrusselsSprouts	Brussel sprouts	TRUE	TRUE

Table 2. Food frequency questionnaire variables included in each food group and descriptions

Food group	FFQ Variable Name	FFQ Description	Included FFQ1	Included FFQ2
Cruciferous vegetables	Cabbage	Cabbage	TRUE	TRUE
Cruciferous vegetables	Cauliflower	Cauliflower	TRUE	TRUE
Cruciferous vegetables	Coleslaw	Coleslaw	TRUE	TRUE
Dairy desserts	DairyDesserts	Dairy desserts (small pot) e.g. chocolate mousse, cream caramels	TRUE	TRUE
Eggs	Eggs	Eggs as boiled, fried, scrambled, etc. (one)	TRUE	TRUE
Fried fish	MFFishFingers	Fish fingers, fish cakes & breaded fish	TRUE	TRUE
Fried fish	MFFriedfish	Fried fish in batter, as in fish and chips	TRUE	TRUE
Fried potatoes	ChipsRoastPots	Chips, roast potatoes	FALSE	TRUE
Fried potatoes	Chips	Chips, retail, fried in vegetable oil	TRUE	FALSE
Fried potatoes	PotatoSalad	Potato salad	TRUE	TRUE
Fried potatoes	RoastPotatoes	Old potatoes, roast in blended oil	TRUE	FALSE
Fruit juice	Smoothies	Smoothies (cup)	FALSE	TRUE
Fruit juice	PureFruitJuice	Pure fruit juice (100%) e.g. orange, apple juice (cup)	TRUE	TRUE
Grapes	Grapes	Grapes (handful)	TRUE	TRUE
Green leafy vegetables	GreenSalad	Green salad, lettuce, cucumber, celery	TRUE	TRUE
Green leafy vegetables	Spinach	Spinach	TRUE	TRUE
Green leafy vegetables	Watercress	Watercress	TRUE	TRUE
Herbal tea	FruitTea	Fruit tea (cup)	FALSE	TRUE
Herbal tea	GreenTea	Green tea (cup)	FALSE	TRUE
High fat milk	ChannelIslandMilk	Channel Islands milk	TRUE	TRUE
High fat milk	EvaporatedMilk	Evaporated milk	TRUE	TRUE
High fat milk	FullMilk	Full cream milk	TRUE	TRUE
High fat milk	CoffeeWhitener	Coffee whitener, e.g. Coffee-mate (teaspoon)	TRUE	TRUE
High fat salad dressing	FrenchDressing	French dressing (tablespoon)	TRUE	TRUE
High fat salad dressing	Mayo	Full fat salad cream, mayonnaise (tablespoon)	TRUE	TRUE
High fat salad dressing	Otherdressing	Other salad dressing (tablespoon)	TRUE	TRUE
High fibre breakfast cereals	HighFibreCereal	High Fibre cereals e.g. Branflakes, All Bran, Fruit and Fibre	FALSE	TRUE
High fibre breakfast cereals	AllBran	All Bran	TRUE	FALSE
High fibre breakfast cereals	Branflakes	Branflakes	TRUE	FALSE
High fibre breakfast cereals	Cheerios	Cheerios	TRUE	FALSE
High fibre breakfast cereals	FruitnFibre	Fruit n' fibre	TRUE	FALSE
High fibre breakfast cereals	Grapenuts	Grape nuts	TRUE	FALSE
High fibre breakfast cereals	Muesli	Muesli	TRUE	TRUE
High fibre breakfast cereals	OatBasedCereal	Oat based	TRUE	FALSE

Table 2. Food frequency questionnaire variables included in each food group and descriptions

Food group	FFQ Variable Name	FFQ Description	Included FFQ1	Included FFQ2
High fibre breakfast cereals	OtherCereal	Cereal, other	TRUE	FALSE
High fibre breakfast cereals	ShreddedWheat	Shredded wheat	TRUE	FALSE
High fibre breakfast cereals	Shreddies	Shreddies	TRUE	FALSE
High fibre breakfast cereals	Start	Start	TRUE	FALSE
High fibre breakfast cereals	SultanaBran	Sultana Bran	TRUE	FALSE
High fibre breakfast cereals	WeetabixType	Weetabix type	TRUE	FALSE
High fibre breakfast cereals	WheatFlakes	Wheat flakes	TRUE	FALSE
High sugar drinks	FruitSquash	Fruit squash or cordial (cup)	TRUE	TRUE
High sugar drinks	FizzySoftDrinks	Fizzy soft drinks, e.g. Coca Cola, lemonade (cup)	TRUE	TRUE
Ice cream	IceCream	Ice cream, choc ices	TRUE	TRUE
Lasagne	Lasagne	Lasagne, moussaka	TRUE	TRUE
Legumes	Beansprouts	Beansprouts	TRUE	TRUE
Legumes	DriedLentils	Pulses e.g. lentils, beans, peas	TRUE	TRUE
Legumes	GreenBeans	Green beans, broad beans, runner beans	TRUE	TRUE
Legumes	Peas	Peas	TRUE	TRUE
Low fat milk	DriedMilk	Dried milk	TRUE	TRUE
Low fat milk	SemiSkimmedMilk	Semi-skimmed milk	TRUE	TRUE
Low fat milk	SkimmedMilk	Skimmed milk	TRUE	TRUE
Low fat salad dressing	Lowcalsaladcream	Low calorie, low fat salad cream (tablespoon)	TRUE	TRUE
Low fat spread	LowFatSpread	Low fat spread, e.g. Outline, Gold (teaspoon)	TRUE	TRUE
Low fat spread	VLowFatSpread	Very low fat spread (teaspoon) e.g. Diet Flora	TRUE	TRUE
Low fibre breakfast cereals	BreakfastCereal	Breakfast cereal e.g. Cornflakes, Rice Krispies	FALSE	TRUE
Low fibre breakfast cereals	CocoPops	Coco pops	TRUE	FALSE
Low fibre breakfast cereals	Cornflakes	Corn Flakes	TRUE	FALSE
Low fibre breakfast cereals	CrunchynutCornflakes	Crunchynut cornflakes	TRUE	FALSE
Low fibre breakfast cereals	Frosties	Sugar topped cereals e.g. Frosties	TRUE	TRUE
Low fibre breakfast cereals	JustRightType	Just right type	TRUE	FALSE
Low fibre breakfast cereals	RiceKrispies	Rice Krispies	TRUE	FALSE
Low fibre breakfast cereals	SpecialK	Special K	TRUE	FALSE
Low fibre breakfast cereals	SugarPuffType	Sugar puffs	TRUE	FALSE
Low sugar soda	Dietfizzy	Low calorie or diet fizzy soft drinks (cup)	TRUE	TRUE
Malt drinks	Horlicks	Horlicks, Ovaltine (cup)	TRUE	TRUE
Margarine	CholLowerSpread	Cholesterol lowering fat spreads e.g. Benecol (teaspoon)	FALSE	TRUE
Margarine	OliveOilSpread	Olive oil spread (teaspoon)	FALSE	TRUE
Margarine	BlockMarg	Block margarine, e.g. Stork, Krona (teaspoon)	TRUE	TRUE

Table 2. Food frequency questionnaire variables included in each food group and descriptions

Food group	FFQ Variable Name	FFQ Description	Included FFQ1	Included FFQ2
Margarine	OtherSoftMarg	Other soft margarine, dairy spreads, e.g. Blue Band, Clover (teaspoon)	TRUE	TRUE
Marrow	Marrow	Marrow, courgettes	TRUE	TRUE
Meat	MFBeef	Beef: roast, steak, mince, stew or casserole	TRUE	TRUE
Meat	MFLamb	Lamb: roast, chops or stew	TRUE	TRUE
Meat	MFPork	Pork: roast, chops or stew	TRUE	TRUE
Melon	Melon	Melon (1 slice)	TRUE	TRUE
Mushrooms	Mushrooms	Mushrooms	TRUE	TRUE
Nuts	NutsSalted	Salted nuts e.g. peanuts, cashews (handful)	FALSE	TRUE
Nuts	NutsUnsalted	Unsalted nuts, e.g. brazil, walnuts (handful)	FALSE	TRUE
Nuts	Seeds	Seeds e.g. Sunflower, pumpkin (tablespoon)	FALSE	TRUE
Nuts	Nuts	Mixed nuts	TRUE	FALSE
Nuts	PeanutButter	Peanut butter (teaspoon)	TRUE	TRUE
Oily fish	MFOilyfish	Oily fish, fresh or canned, e.g. tuna, mackerel, kippers, salmon, sardines, herring	TRUE	TRUE
Other fish/seafood	MFFishroe	Fish roe, taramasalata	TRUE	TRUE
Other fish/seafood	MFShellfish	Shellfish, e.g. crab, prawns, mussels	TRUE	TRUE
Other fish/seafood	MFWhitefish	Other white fish, fresh or frozen, e.g. cod, plaice, sole, haddock, halibut	TRUE	TRUE
Peaches	Peaches	Peaches, plums, apricots (1 fruit)	TRUE	TRUE
Pizza	Pizza	Pizza (one slice)	TRUE	TRUE
Pizza	Quiche	Quiche (slice)	TRUE	TRUE
Polyunsaturated margarine	PufaMarg	Polyunsaturated margarine, e.g. Flora, sunflower (teaspoon)	TRUE	TRUE
Porridge	Porridge	Porridge, Readybreak, oats	TRUE	TRUE
Poultry	MFPoultry	Chicken or other poultry e.g. turkey	TRUE	TRUE
Processed fruit	DriedFruit	Dried fruit, e.g. raisins, prunes (heaped tablespoon)	TRUE	TRUE
Processed fruit	TinnedFruit	Tinned fruit (handful)	TRUE	TRUE
Processed meats	MFBacon	Bacon or gammon	TRUE	TRUE
Processed meats	MFCornedBeef	Corned Beef, Spam, luncheon meats	TRUE	TRUE
Processed meats	MFHam	Ham, cured meats & chorizo	TRUE	TRUE
Processed meats	MFLiver	Liver, liver pate, liver sausage	TRUE	TRUE
Processed meats	MFSausages	Sausages	TRUE	TRUE
Root vegetables	Beetroot	Beetroot	TRUE	TRUE
Root vegetables	Parsnips	Parsnips, turnips, swedes	TRUE	TRUE
Root vegetables	Carrots	Carrots	TRUE	TRUE
Savoury pies	MFPies	Savoury pies, e.g. meat pie, pork pie, pasties, steak & kidney pie, sausage rolls	TRUE	TRUE
Savoury snacks	CreamCrackers	Cream crackers, savoury biscuits	TRUE	TRUE
Savoury snacks	Crisps	Crisps or other packet snacks, e.g. Wotsits (one packet)	TRUE	TRUE

Table 2. Food frequency questionnaire variables included in each food group and descriptions

Food group	FFQ Variable Name	FFQ Description	Included FFQ1	Included FFQ2
Seasonings	Marmite	Marmite, Bovril (teaspoon)	TRUE	TRUE
Seasonings	PicklesChutney	Pickles, chutney (tablespoon)	TRUE	TRUE
Seasonings	Sauces	Sauces, e.g. white sauce, cheese sauce, gravy (tablespoon)	TRUE	TRUE
Seasonings	TomatoKetchup	Tomato ketchup (tablespoon)	TRUE	TRUE
Soup	MeatSoup	Meat soups (bowl) (to include meat and vegetable soups)	TRUE	TRUE
Soup	VegSoup	Vegetable soups (bowl)	TRUE	TRUE
Soy foods	Tofu	Meat substitutes e.g. tofu, soyameat, textured vegetable protein, vegeburger	TRUE	TRUE
Soy/other milk	GoatsMilk	Goats' milk	TRUE	TRUE
Soy/other milk	RiceMilk	Rice milk	TRUE	TRUE
Soy/other milk	SoyaMilk	Soya milk	TRUE	TRUE
Spirits/liquor	Liqueurs	Port, sherry, vermouth, liqueurs (pub measure)	TRUE	TRUE
Spirits/liquor	Spirits	Spirits, e.g. gin, brandy, whisky, vodka (pub measure)	TRUE	TRUE
Sweet baked products	BiscuitsRedFat	Reduced fat biscuits e.g. Go Ahead, Highlights (one small packet or one small bar/biscuit)	FALSE	TRUE
Sweet baked products	Biscuitschoc	Sweet biscuits, chocolate, e.g. digestive (one)	TRUE	TRUE
Sweet baked products	Biscuitsplain	Sweet biscuits, plain, e.g. Nice, ginger (one)	TRUE	TRUE
Sweet baked products	Bunshome	Buns, pastries e.g. scones, flapjacks, croissants, doughnuts, home baked	TRUE	TRUE
Sweet baked products	Cakeshome	Cakes e.g. fruit, sponge, home baked	TRUE	TRUE
Sweet baked products	Cakesready	Cakes e.g. fruit, sponge, ready made	TRUE	TRUE
Sweet baked products	Fruitpiehome	Fruit pies, tarts, crumbles, home baked	TRUE	TRUE
Sweet baked products	Fruitpiesready	Fruit pies, tarts, crumbles, ready made	TRUE	TRUE
Sweet baked products	MilkPuddings	Milk puddings e.g. rice, custard, trifle	TRUE	TRUE
Sweet baked products	SpongePudhome	Sponge puddings, home baked	TRUE	TRUE
Sweet baked products	SpongePudready	Sponge puddings, ready made	TRUE	TRUE
Sweet peppers	SweetPeppers	Sweet peppers	TRUE	TRUE
Sweetcorn	Sweetcorn	Sweetcorn	TRUE	TRUE
Tomatoes	Tomatoes	Tomatoes	TRUE	TRUE
White/brown bread, refined grains	NaanPoppdmTort	Naan, poppadoms, flour tortillas	FALSE	TRUE
White/brown bread, refined grains	BrownBread	Brown bread/rolls	TRUE	TRUE
White/brown bread, refined grains	WhiteBread	White bread/rolls	TRUE	TRUE
White/brown bread, refined grains	Whitepasta	White or green pasta, e.g. spaghetti, macaroni, noodles	TRUE	TRUE

Table 2. Food frequency questionnaire variables included in each food group and descriptions

Food group	FFQ Variable Name	FFQ Description	Included FFQ1	Included FFQ2
White/brown bread, refined grains	WhiteRice	White rice	TRUE	TRUE
Wholemeal bread/grains	BrownRice	Brown rice	TRUE	TRUE
Wholemeal bread/grains	WholemealBread	Wholemeal & granary bread/rolls	TRUE	TRUE
Wholemeal bread/grains	WholemealPasta	Wholemeal pasta	TRUE	TRUE
Wine	RedWine	Red wine (small glass)	FALSE	TRUE
Wine	WhiteWine	White wine (small glass)	FALSE	TRUE
Wine	Wine	White wine, red wine	TRUE	FALSE
Wine	Wine	White wine, red wine	TRUE	FALSE
Yoghurt	DairyFFYog	Full fat or Greek yoghurt (small pot)	TRUE	TRUE
Yoghurt	DairyLFYog	Low fat yoghurt, fromage frais (small pot)	TRUE	TRUE

Notes: Details of each food frequency questionnaire (FFQ) variables included in each food group analysis are shown. FFQs completed before 2007 (FFQ1) included slightly different variables than FFQs completed during 2007 (FFQ2), therefore the right hand columns indicate if a particular variable was included in the final (TRUE) or not (FALSE).

Figure 1. Scree plot for the first 20 principal components, adapted from (Teucher et al., 2007)

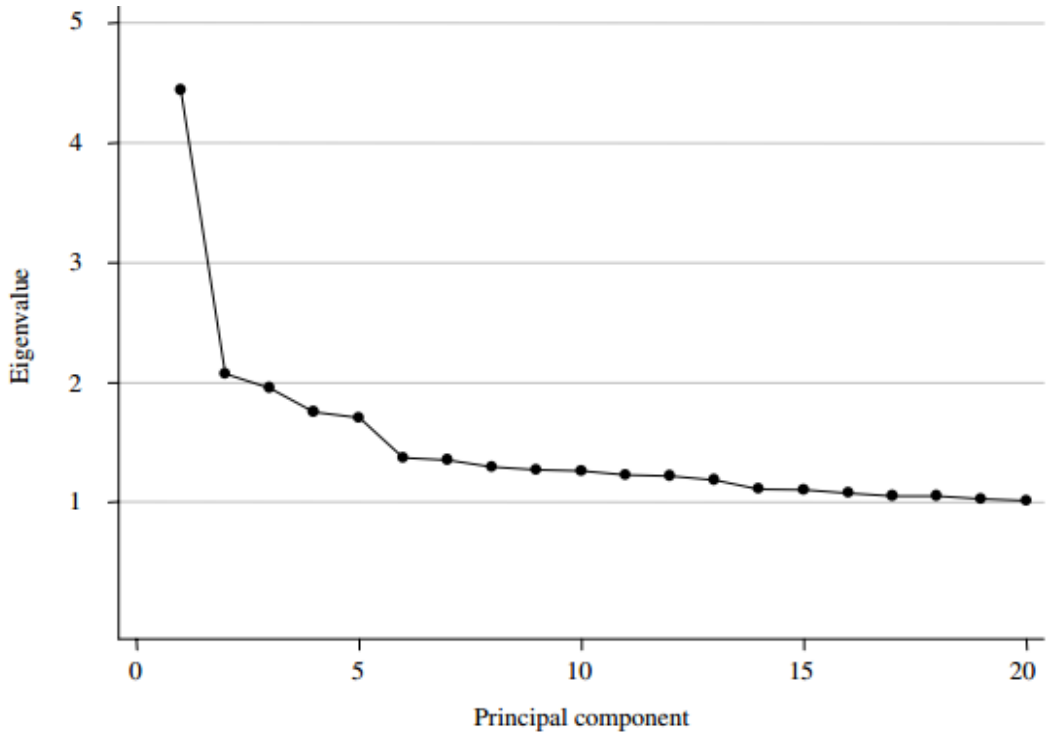


Table 3. Factor loadings for the five principal components (modified from (Teucher et al., 2007))

Food	Fruit and Vegetable	High Alcohol	Traditional English	Dieting	Low Meat
Baked beans	–.01 (–.04,.01)	–.07 (–.18,.10)	.17 (–.01,.25)	.10 (–.19,.36)	.31 (–.16,.37)
Beefburgers	–.14 (–.16,–.13)	.13 (–.11,.23)	.18 (–.02,.25)	.15 (–.02,.23)	.08 (–.18,.19)
Beer	–.03 (–.05,–.01)	.22 (.07,.26)	.00 (–.18,.17)	.09 (–.03,.16)	.04 (–.12,.14)
Berries	.16 (.14,.19)	–.05 (–.11,.04)	–.05 (–.11,.04)	.00 (–.10,.09)	–.05 (–.13,.06)
Butter	–.07 (–.09,–.05)	.11 (.00,.19)	–.03 (–.16,.13)	–.35 (–.40,–.03)	–.20 (–.40,.37)
Citrus fruit	.19 (.16,.21)	–.11 (–.16,–.01)	–.01 (–.12,.09)	.08 (–.11,.16)	–.07 (–.15,.06)
Coffee	–.01 (–.04,.01)	.17 (.04,.21)	–.01 (–.17,.14)	.13 (–.20,.23)	–.16 (–.25,.06)
Cooked potatoes	.03 (.00,.05)	–.19 (–.27,.00)	.19 (–.02,.28)	–.08 (–.17,.08)	–.02 (–.15,.15)
Crisp bread	.11 (.09,.14)	–.06 (–.11,.02)	–.03 (–.10,.05)	.03 (–.07,.10)	.02 (–.08,.11)
Dairy products; high fat	–.07 (–.09,–.06)	.14 (.00,.20)	–.13 (–.21,.02)	–.12 (–.20,.02)	–.07 (–.18,.16)
Dairy products; low fat	.15 (.13,.17)	–.29 (–.33,–.06)	–.05 (–.24,.21)	.25 (–.09,.29)	–.02 (–.27,.22)
Drinks; other	.04 (.02,.06)	–.07 (–.14,.07)	–.10 (–.17,.02)	.17 (–.01,.23)	.11 (–.19,.23)
Eggs	.00 (–.02,.02)	.04 (–.08,.12)	.11 (.01,.17)	–.05 (–.14,.07)	–.06 (–.16,.11)
Fried fish	–.11 (–.13,–.09)	–.02 (–.18,.13)	.21 (.08,.24)	–.01 (–.11,.10)	.03 (–.07,.13)
Fried potatoes	–.21 (–.22,–.19)	.09 (–.17,.23)	.22 (.04,.27)	.01 (–.12,.17)	.12 (–.05,.19)
Fruit juice	.09 (.07,.12)	.06 (–.03,.12)	–.07 (–.14,.03)	.08 (–.08,.13)	–.04 (–.13,.07)
High fibre breakfast cereals	.06 (.04,.09)	–.23 (–.29,.04)	–.14 (–.26,.12)	.13 (–.17,.23)	–.13 (–.27,.08)
Lasagne	.01 (–.01,.03)	.13 (.02,.18)	.00 (–.14,.13)	.25 (–.00,.31)	.13 (–.27,.29)
Legumes	.19 (.16,.21)	.03 (–.24,.23)	.29 (.10,.33)	–.16 (–.27,.18)	.13 (–.09,.28)
Low fibre breakfast cereals	–.07 (–.09,–.05)	–.06 (–.11,.01)	.05 (–.04,.11)	–.05 (–.12,.10)	.07 (–.03,.14)
Low fat spread	.02 (–.01,.04)	–.10 (–.14,–.01)	.02 (–.09,.12)	.15 (–.01,.22)	.11 (–.18,.22)
Margarine	–.09 (–.11,–.07)	–.01 (–.08,.07)	.07 (–.01,.12)	.00 (–.12,.16)	.13 (–.04,.20)
Meat	–.10 (–.12,–.08)	–.02 (–.27,.21)	.34 (.11,.41)	.05 (–.35,.31)	–.32 (–.39,.02)
Nuts	.06 (.02,.09)	0.14 (–.02,.24)	–.15 (–.25,.01)	–.12 (–.20,.11)	.07 (–.09,.18)
Oily fish	.15 (.12,.19)	0.01 (–.04,.07)	–.01 (–.09,.07)	.02 (–.16,.13)	–.12 (–.17,.02)
Other fish and seafood	.18 (.16,.21)	0.07 (–.03,.13)	.07 (–.06,.18)	–.04 (–.25,.19)	–.20 (–.26,.07)
Other fruit	.27 (.25,.29)	–.21 (–.26,.00)	–.07 (–.21,.13)	.09 (–.06,.14)	–.01 (–.13,.10)
Pizza	–.06 (–.08,–.04)	0.15 (.02,.22)	–.10 (–.21,.07)	.18 (–.09,.31)	.22 (–.20,.32)
Polyunsaturated margarine	–.02 (–.04,.00)	–.08 (–.14,.03)	–.06 (–.15,.06)	.07 (–.10,.21)	.18 (–.10,.25)
Porridge	.08 (.06,.10)	–.13 (–.17,–.03)	.05 (–.08,.15)	–.14 (–.20,.05)	.00 (–.15,.17)
Poultry	.05 (.03,.07)	–.02 (–.13,.09)	.14 (.01,.25)	.17 (–.30,.32)	–.25 (–.33,.04)
Processed meats	–.10 (–.12,–.08)	.02 (–.25,.24)	.32 (.10,.38)	.13 (–.23,.28)	–.16 (–.27,.06)
Salad dressing high fat	.12 (.09,.14)	.30 (.08,.34)	–.11 (–.31,.13)	–.02 (–.10,.08)	–.03 (–.10,.07)
Salad dressing low fat	.14 (.10,.17)	–.02 (–.13,.11)	.10 (–.03,.19)	.18 (–.03,.27)	.13 (–.20,.25)
Savoury pies	–.19 (–.21,–.17)	.03 (–.20,.19)	.24 (.09,.27)	.00 (–.10,.09)	–.01 (–.09,.09)
Savoury snacks	–.10 (–.12,–.08)	.12 (.03,.17)	–.03 (–.12,.08)	.04 (–.07,.13)	.09 (–.07,.15)
Seasonings	.02 (.00,.05)	.05 (–.05,.11)	.07 (–.04,.14)	.02 (–.15,.23)	.18 (–.06,.25)
Soda; high sugar	–.08 (–.10,–.06)	.09 (–.01,.15)	.04 (–.07,.12)	.08 (–.05,.19)	.11 (–.12,.20)
Soda; low sugar	.02 (–.01,.04)	.06 (–.05,.14)	.04 (–.10,.18)	.33 (–.05,.36)	.08 (–.34,.33)
Soup	.12 (.10,.14)	–.03 (–.10,.03)	.03 (–.05,.11)	.05 (–.17,.18)	–.13 (–.20,.02)
Soy and other milk	.05 (.03,.08)	.00 (–.06,.08)	–.03 (–.11,.05)	–.15 (–.22,.12)	.08 (–.10,.21)
Soy foods	.11 (.07,.14)	.06 (–.06,.17)	–.09 (–.24,.05)	–.15 (–.37,.40)	.38 (.03,.43)
Spirits and liquor	.00 (–.02,.02)	.20 (.06,.23)	.01 (–.17,.16)	.10 (–.18,.20)	–.13 (–.21,.04)

Table 3. Factor loadings for the five principal components (modified from (Teucher et al., 2007))

Food	Fruit and Vegetable	High Alcohol	Traditional English	Dieting	Low Meat
Sweet baked	–.16 (–.17,–.14)	–.22 (–.26,.00)	–.10 (–.25,.14)	–.22 (–.27,.10)	.00 (–.20,.24)
Sweets and sweet condiments	–.17 (–.19,–.15)	–.02 (–.08,.05)	–.01 (–.13,.09)	–.26 (–.29,.17)	.09 (–.19,.28)
Tea	–.03 (–.05,–.01)	–.23 (–.28,–.05)	.08 (–.13,.26)	–.24 (–.29,.12)	.02 (–.22,.27)
Vegetables; allium	.23 (.21,.25)	.27 (.04,.30)	.06 (–.18,.23)	–.13 (–.18,.07)	.02 (–.13,.18)
Vegetables; cruciferous	.21 (.17,.24)	–.03 (–.32,.26)	.36 (.14,.39)	–.08 (–.20,.15)	.07 (–.08,.19)
Vegetables; green leafy	.30 (.28,.32)	.12 (–.04,.17)	.07 (–.06,.15)	–.06 (–.12,.10)	.07 (–.04,.13)
Vegetables; other	.32 (.30,.33)	.18 (–.09,.26)	.15 (–.05,.25)	–.13 (–.20,.15)	.12 (–.07,.19)
Vegetables; yellow	.30 (.28,.32)	–.03 (–.17,.13)	.18 (.06,.23)	–.09 (–.15,.11)	.06 (–.07,.15)
White and brown bread, refined grains	–.11 (–.13,–.09)	.04 (–.02,.09)	–.03 (–.10,.05)	–.02 (–.14,.17)	.15 (–.04,.22)
Wholemeal bread and grains	.15 (.12,.17)	–.11 (–.22,.12)	–.18 (–.23,.00)	–.01 (–.10,.08)	–.01 (–.10,.08)
Wine	.08 (.06,.10)	.33 (.07,.37)	–.14 (–.35,.14)	.03 (–.25,.21)	–.22 (–.27,.02)

Table 4. Items and gram amounts per frequency category for the Mediterranean diet score calculation

Fruit & nuts		Vegetable		Meat		Fish		Dairy		Cereal[1]		Legume	
Item	***g***	***Item***	***g***	***Item***	***g***	***Item***	***g***	***Item***	***g***	***Item***	***g***	***Item***	***g***
Apples	100	Avocado	75	Bacon	25	Shellfish	60	Channel Islands milk	585	Brown bread	36	Baked beans	135
Bananas	100	Beansprouts	20	Beef	128	Oilyfish	96	Cottage cheese	55	Brown rice	180	Lentils	70
Dried fruit	30	Beetroot	40	Burgers	78	Fried fish	180	Dairy cheese	40	Cream crackers	7		
Tinned fruit	100	Broccoli	85	Corned beef	30	Fish roe	45	Dairy deserts	125	Crispbread	10		
Grapefruit	80	Cabbage	95	Ham	23	Fish fingers	50	Low fat yogurt	125	Frosties	40		
Grapes	80	Brussel sprouts	90	Lamb	94	Whitefish	120	Double cream	30	Lasagne	420		
Melon	200	Carrots	60	Liver	40			Single cream	15	White pasta	230		
Oranges	120	Cauliflower	90	Meat pies	124			Dried milk	585	Porridge	160		
Peaches	70	Coleslaw	45	Pork	128			Eggs	50	Muesli	60		
Pears	170	Garlic	5	Poultry	125			Full milk	585	Wholemeal pasta	230		
Strawberries	100	Green beans	90	Sausages	40			Goats milk	585	White bread	36		
Mixed nuts[2]	40	Green salad	30					Rice milk	585	White rice	180		
Nuts unsalted[2]	40	Leeks	75					Semi-skimmed milk	585	Wholemeal bread	36		
Nuts salted[2]	40	Marrow	90					Skimmed milk	585				
		Mushrooms	56					Soya milk	585				
		Onions	60					Evaporated milk	585				
		Peas	70										
		Spinach	90										
		Sweetcorn	85										
		Sweet peppers	80										
		Tomatoes	85										
		Watercress	20										
		Tofu	50										
		Parsnips	65										

Notes: [1]The cereal category for FFQ 1 included multiple cereal types which were omitted from FFQ 2, therefore only items which were included in both questionnaires were quantified.

[2]'Mixed nuts' was quantified on only FFQ 1, 'nuts unsalted' and 'nuts salted' were quantified on FFQ 2. For this analysis, frequencies for these items were combined to make one category.

Table 5. List of metabolites associated with food intake from the Metabolon platform

				Discovery		Discordant		Meta-analysis		Whole sample
Variable	Metabolite name	Pathway	Super-pathway	Beta(SE)	*P*	Beta(SE)	*P*	Beta(SE)	*P*	R^2
Allium vegetables	X-11858			0.040(0.007)	3.65E-08	0.036(0.015)	1.69E-02	0.039(0.007)	1.42E-09	0.030
Allium vegetables	tryptophan betaine	Tryptophan metabolism	Amino acid	0.041(0.007)	3.82E-08	0.023(0.010)	1.57E-02	0.034(0.006)	4.73E-09	0.025
Allium vegetables	X-02269			0.024(0.005)	2.82E-07	0.019(0.010)	4.75E-02	0.023(0.004)	3.55E-08	0.010
Apples/pears	threitol	Nucleotide sugars, pentose metabolism	Carbo-hydrate	0.034(0.004)	2.50E-17	0.030(0.007)	4.05E-05	0.033(0.003)	1.69E-21	0.031
Apples/pears	X-11315			0.035(0.004)	9.14E-15	0.036(0.008)	6.10E-06	0.035(0.004)	9.63E-20	0.029
Apples/pears	X-11372			-0.024(0.004)	1.97E-09	-0.021(0.007)	3.40E-03	-0.023(0.003)	2.02E-11	0.014
Apples/pears	indolepropionate	Tryptophan metabolism	Amino acid	0.027(0.005)	4.50E-07	0.024(0.008)	1.71E-03	0.026(0.004)	2.39E-09	0.016
Apples/pears	3-phenylpropionate (hydrocinnamate)	Phenylalanine & tyrosine metabolism	Amino acid	0.025(0.005)	3.76E-07	0.023(0.009)	1.21E-02	0.024(0.004)	1.24E-08	0.015
Apples/pears	X-09789			0.021(0.004)	2.73E-07	0.016(0.007)	2.58E-02	0.020(0.004)	2.22E-08	0.010
Avocado	X-11469			0.204(0.027)	3.59E-14	0.100(0.027)	3.31E-04	0.153(0.019)	9.36E-16	0.023
Avocado	X-02269			0.193(0.028)	3.88E-12	0.107(0.026)	6.44E-05	0.147(0.019)	6.33E-15	0.021
Avocado	3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF)	Fatty acid, dicarboxylate	Lipid	0.185(0.027)	8.83E-12	0.116(0.029)	1.26E-04	0.153(0.020)	1.03E-14	0.022
Avocado	1-docosahexaenoylglycerophosphocholine*	Lysolipid	Lipid	0.125(0.024)	2.54E-07	0.094(0.028)	1.06E-03	0.112(0.018)	1.03E-09	0.009
Avocado	eicosapentaenoate (EPA; 20:5n3)	Essential fatty acid	Lipid	0.163(0.026)	3.82E-10	0.056(0.025)	2.56E-02	0.108(0.018)	2.01E-09	0.013
Bananas	indolepropionate	Tryptophan metabolism	Amino acid	0.038(0.006)	5.61E-11	0.022(0.010)	3.37E-02	0.034(0.005)	1.05E-11	0.016
Beef burgers	trans-4-hydroxyproline	Urea cycle; arginine-, proline-, metabolism	Amino acid	0.428(0.085)	5.90E-07	0.257(0.104)	1.48E-02	0.360(0.066)	5.34E-08	0.009
Berries	X-11315			0.124(0.018)	2.28E-12	0.072(0.021)	8.29E-04	0.103(0.013)	2.70E-14	0.021
Black tea	X-14473			-0.025(0.001)	8.51E-60	-0.020(0.003)	5.16E-09	-0.024(0.001)	1.36E-72	0.110
Black tea	quinate	Food component, Plant	Xenobiotics	-0.019(0.002)	9.74E-34	-0.019(0.003)	1.61E-09	-0.019(0.001)	5.58E-44	0.066
Black tea	X-12816			-0.019(0.002)	4.68E-23	-0.021(0.004)	2.59E-07	-0.020(0.002)	3.57E-30	0.062

Table 5. List of metabolites associated with food intake from the Metabolon platform

Variable	Metabolite name	Pathway	Super-pathway	Discovery		Discordant		Meta-analysis		Whole sample
				Beta(SE)	*P*	Beta(SE)	*P*	Beta(SE)	*P*	R^2
Black tea	X-14374			-0.013(0.001)	8.48E-21	-0.011(0.003)	2.12E-03	-0.013(0.001)	2.50E-23	0.030
Black tea	X-12039--3-hydroxypyridine sulfate	Chemical	Xenobiotics	-0.014(0.002)	2.23E-16	-0.013(0.004)	4.61E-04	-0.014(0.002)	1.60E-19	0.035
Black tea	X-05426			-0.010(0.002)	8.54E-12	-0.008(0.003)	1.66E-02	-0.010(0.001)	3.62E-13	0.019
Black tea	cyclo(leu-pro)	Dipeptide	Peptide	-0.012(0.002)	5.64E-11	-0.010(0.004)	2.26E-02	-0.012(0.002)	2.89E-12	0.026
Black tea	X-12230			-0.009(0.002)	3.07E-08	-0.010(0.003)	4.67E-03	-0.009(0.001)	3.98E-10	0.015
Black tea	X-13741--3-methyl catechol sulfate 1	Benzoate metabolism	Xenobiotics	-0.010(0.002)	6.29E-08	-0.011(0.004)	4.92E-03	-0.010(0.002)	8.73E-10	0.017
Butter	X-13431--nonanoylcarnitine*	Carnitine metabolism	Lipid	0.028(0.004)	9.12E-15	0.020(0.006)	1.97E-03	0.026(0.003)	6.48E-17	0.026
Butter	X-02249			0.027(0.003)	6.97E-15	0.019(0.007)	1.06E-02	0.026(0.003)	2.31E-16	0.023
Butter	X-10510			0.023(0.003)	1.85E-11	0.019(0.005)	1.50E-04	0.022(0.003)	8.28E-15	0.019
Butter	15-methylpalmitate (isobar with 2-methylpalmitate)	Fatty acid, branched	Lipid	0.027(0.004)	6.27E-11	0.019(0.008)	1.38E-02	0.025(0.004)	3.09E-12	0.026
Butter	10-undecenoate (11:1n1)	Medium chain fatty acid	Lipid	0.023(0.004)	5.99E-10	0.014(0.005)	4.11E-03	0.020(0.003)	2.35E-11	0.018
Butter	myristate (14:0)	Long chain fatty acid	Lipid	0.019(0.004)	5.57E-08	0.019(0.006)	4.34E-03	0.019(0.003)	6.68E-10	0.014
Butter	X-08402			0.018(0.003)	1.85E-08	0.013(0.005)	1.10E-02	0.017(0.003)	7.17E-10	0.012
Butter	10-nonadecenoate (19:1n9)	Long chain fatty acid	Lipid	0.017(0.004)	9.50E-07	0.019(0.006)	8.91E-04	0.018(0.003)	2.43E-09	0.013
Butter	pentadecanoate (15:0)	Long chain fatty acid	Lipid	0.019(0.004)	8.48E-08	0.015(0.006)	1.03E-02	0.018(0.003)	2.77E-09	0.013
Chocolate	theobromine	Xanthine metabolism	Xenobiotics	0.024(0.005)	3.43E-07	0.023(0.005)	1.83E-05	0.024(0.003)	1.34E-11	0.020
Chocolate	7-methylxanthine	Xanthine metabolism	Xenobiotics	0.029(0.005)	2.48E-08	0.012(0.006)	3.31E-02	0.022(0.004)	2.00E-08	0.020
Citrus fruit	stachydrine	Food component, Plant	Xenobiotics	0.058(0.006)	6.43E-24	0.036(0.008)	4.05E-05	0.051(0.005)	2.07E-27	0.051
Citrus fruit	glycerate	Glycolysis, gluconeogenesis, pyruvate metabolism	Carbohydrate	0.027(0.005)	1.15E-07	0.031(0.008)	1.39E-04	0.028(0.004)	4.57E-11	0.014
Citrus fruit	X-11315			0.029(0.005)	4.72E-08	0.021(0.007)	3.30E-03	0.026(0.004)	6.50E-10	0.014
Coffee	X-14473			0.041(0.001)	6.89E-139	0.027(0.003)	1.38E-19	0.038(0.001)	6.12E-187	0.246
Coffee	quinate	Food component, Plant	Xenobiotics	0.036(0.001)	6.60E-106	0.024(0.003)	6.84E-15	0.033(0.001)	4.31E-136	0.190

Table 5. List of metabolites associated with food intake from the Metabolon platform

Variable	Metabolite name	Pathway	Super-pathway	Discovery		Discordant		Meta-analysis		Whole sample
				Beta(SE)	*P*	Beta(SE)	*P*	Beta(SE)	*P*	R^2
Coffee	X-14374			0.025(0.002)	1.61E-55	0.019(0.003)	1.78E-10	0.023(0.001)	1.87E-68	0.083
Coffee	X-12039--3-hydroxypyridine sulfate	Chemical	Xenobiotics	0.026(0.002)	6.89E-52	0.020(0.003)	1.72E-09	0.025(0.001)	1.64E-64	0.108
Coffee	X-12816			0.032(0.002)	1.07E-51	0.021(0.004)	7.67E-07	0.030(0.002)	3.22E-61	0.133
Coffee	X-05426			0.019(0.002)	5.06E-34	0.012(0.003)	2.06E-04	0.018(0.001)	6.64E-38	0.058
Coffee	X-12230			0.019(0.002)	1.07E-27	0.015(0.003)	1.99E-06	0.018(0.002)	1.02E-33	0.060
Coffee	X-13741--3-methyl catechol sulfate 1	Benzoate metabolism	Xenobiotics	0.020(0.002)	3.43E-27	0.016(0.004)	1.05E-05	0.019(0.002)	1.25E-32	0.063
Coffee	catechol sulfate	Benzoate metabolism	Xenobiotics	0.015(0.002)	5.39E-22	0.011(0.003)	2.24E-04	0.014(0.001)	3.44E-25	0.033
Coffee	cyclo(leu-pro)	Dipeptide	Peptide	0.014(0.002)	9.25E-14	0.009(0.004)	1.58E-02	0.013(0.002)	5.29E-15	0.030
Coffee	X-12217--O-methyl catechol sulfate	Benzoate metabolism	Xenobiotics	0.012(0.002)	5.43E-12	0.009(0.003)	8.66E-03	0.011(0.002)	1.45E-13	0.022
Coffee	1-methylxanthine	Xanthine metabolism	Xenobiotics	0.010(0.002)	2.82E-07	0.010(0.003)	4.20E-03	0.010(0.002)	3.31E-09	0.016
Confectionary /jam	X-11315			-0.011(0.002)	6.99E-11	-0.006(0.003)	2.77E-02	-0.010(0.001)	1.53E-11	0.017
Confectionary /jam	glycerate	Glycolysis, gluconeogenesis, pyruvate metabolism	Carbo-hydrate	-0.011(0.002)	4.86E-08	-0.008(0.003)	1.30E-02	-0.010(0.002)	2.02E-09	0.014
Confectionary /jam	pipecolate	Lysine metabolism	Amino acid	-0.009(0.002)	4.07E-07	-0.008(0.004)	3.81E-02	-0.008(0.002)	3.94E-08	0.008
Fried fish	X-11372			0.360(0.051)	2.88E-12	0.165(0.049)	8.66E-04	0.258(0.035)	2.81E-13	0.040
Fried fish	3-phenylpropionate (hydrocinnamate)	Phenylalanine & tyrosine metabolism	Amino acid	-0.167(0.032)	2.07E-07	-0.184(0.045)	7.67E-05	-0.172(0.026)	4.12E-11	0.012
Fried fish	X-11315			-0.174(0.035)	6.30E-07	-0.149(0.044)	7.69E-04	-0.164(0.027)	1.54E-09	0.013
Fried fish	scyllo-inositol	Inositol metabolism	Lipid	-0.170(0.031)	6.22E-08	-0.116(0.048)	1.77E-02	-0.154(0.026)	4.44E-09	0.011
Fruit juice	stachydrine	Food component, Plant	Xenobiotics	0.064(0.005)	2.87E-30	0.045(0.008)	1.06E-07	0.058(0.005)	3.26E-37	0.064
Green leafy vegetables	X-11469			0.034(0.005)	9.00E-11	0.025(0.006)	1.24E-04	0.030(0.004)	5.66E-14	0.020
Green leafy vegetables	X-02269			0.033(0.005)	9.43E-10	0.025(0.006)	6.56E-05	0.030(0.004)	2.39E-13	0.018
Green leafy vegetables	X-11315			0.038(0.006)	4.37E-10	0.025(0.007)	5.56E-04	0.033(0.005)	1.71E-12	0.023
Green leafy vegetables	X-11372			-0.028(0.004)	5.23E-10	-0.021(0.009)	1.58E-02	-0.026(0.004)	2.51E-11	0.012

Table 5. List of metabolites associated with food intake from the Metabolon platform

Variable	Metabolite name	Pathway	Super-pathway	Discovery		Discordant		Meta-analysis		Whole sample
				Beta(SE)	*P*	Beta(SE)	*P*	Beta(SE)	*P*	R^2
Green leafy vegetables	3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF)	Fatty acid, dicarboxylate	Lipid	0.028(0.006)	9.29E-07	0.021(0.006)	8.06E-04	0.025(0.004)	3.03E-09	0.013
Green leafy vegetables	1-docosahexaenoylglycerophosphocholine*	Lysolipid	Lipid	0.025(0.005)	9.36E-08	0.017(0.007)	1.87E-02	0.023(0.004)	7.11E-09	0.009
High fat salad dressing	eicosapentaenoate (EPA; 20:5n3)	Essential fatty acid	Lipid	0.048(0.010)	8.34E-07	0.035(0.015)	1.82E-02	0.044(0.008)	5.48E-08	0.010
High fibre breakfast cereals	X-09789			0.046(0.006)	4.90E-14	0.019(0.009)	4.34E-02	0.038(0.005)	7.28E-14	0.021
High fibre breakfast cereals	X-11469			0.032(0.005)	1.54E-10	0.028(0.008)	1.21E-03	0.031(0.004)	5.40E-13	0.013
High fibre breakfast cereals	X-02269			0.031(0.005)	7.68E-10	0.028(0.009)	1.65E-03	0.030(0.004)	3.63E-12	0.012
High fibre breakfast cereals	pyridoxate	Vitamin B6 metabolism	Cofactors and vitamins	0.040(0.006)	9.41E-11	0.020(0.010)	3.54E-02	0.035(0.005)	3.52E-11	0.019
High fibre breakfast cereals	X-11315			0.031(0.006)	8.40E-08	0.027(0.010)	1.12E-02	0.030(0.005)	2.75E-09	0.011
Low fat mik	X-21365 [trimethyl-N-aminovalerate]	Carnitine metabolism	Lipid	0.085(0.008)	3.30E-25	0.045(0.015)	2.55E-03	0.076(0.007)	9.36E-27	0.018
Low fat mik	X-12798			0.071(0.009)	4.88E-15	0.036(0.016)	2.12E-02	0.062(0.008)	1.24E-15	0.024
Meat	trans-4-hydroxyproline	Urea cycle; arginine-, proline-, metabolism	Amino acid	0.077(0.010)	7.70E-14	0.070(0.017)	6.28E-05	0.075(0.009)	1.08E-17	0.020
Meat	creatine	Creatine metabolism	Amino acid	0.070(0.011)	5.01E-11	0.041(0.019)	3.12E-02	0.063(0.009)	8.24E-12	0.016
Meat	pyroglutamine*	Glutamate metabolism	Amino acid	-0.066(0.011)	4.33E-09	-0.049(0.020)	1.46E-02	-0.062(0.010)	2.10E-10	0.016
Meat	X-09789			-0.052(0.011)	1.11E-06	-0.053(0.020)	7.66E-03	-0.052(0.009)	2.39E-08	0.009
Mushrooms	ergothioneine	Food component, Plant	Xenobiotics	0.189(0.023)	6.45E-16	0.165(0.033)	9.28E-07	0.181(0.019)	5.93E-22	0.107
Nuts	X-11315			0.060(0.006)	6.48E-21	0.042(0.009)	1.81E-05	0.054(0.005)	3.75E-25	0.037
Oily fish	docosahexaenoate (DHA; 22:6n3)	Essential fatty acid	Lipid	0.244(0.017)	1.72E-45	0.087(0.019)	9.66E-06	0.177(0.013)	2.09E-44	0.064
Oily fish	X-11469			0.214(0.017)	6.59E-34	0.117(0.022)	1.89E-07	0.176(0.013)	5.87E-39	0.062

Table 5. List of metabolites associated with food intake from the Metabolon platform

Variable	Metabolite name	Pathway	Super-pathway	Discovery		Discordant		Meta-analysis		Whole sample
				Beta(SE)	*P*	Beta(SE)	*P*	Beta(SE)	*P*	R^2
Oily fish	3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF)	Fatty acid, dicarboxylate	Lipid	0.204(0.016)	3.52E-34	0.112(0.022)	9.22E-07	0.172(0.013)	8.57E-39	0.054
Oily fish	X-02269			0.208(0.017)	1.44E-32	0.122(0.022)	7.36E-08	0.175(0.013)	1.44E-38	0.058
Oily fish	eicosapentaenoate (EPA; 20:5n3)	Essential fatty acid	Lipid	0.212(0.017)	1.95E-33	0.085(0.024)	5.01E-04	0.169(0.014)	1.57E-33	0.052
Oily fish	1-docosahexaenoylglycerophosphocholine*	Lysolipid	Lipid	0.145(0.017)	7.40E-17	0.068(0.032)	3.36E-02	0.128(0.015)	3.33E-17	0.025
Oily fish	X-11315			0.138(0.018)	6.44E-15	0.054(0.022)	1.68E-02	0.106(0.014)	1.62E-14	0.020
Oily fish	1-oleoylglycerophosphoethanolamine	Lysolipid	Lipid	-0.115(0.018)	2.65E-10	-0.074(0.026)	4.23E-03	-0.101(0.015)	6.89E-12	0.012
Oily fish	1-eicosatrienoylglycerophosphocholine*	Lysolipid	Lipid	-0.099(0.017)	9.17E-09	-0.087(0.024)	4.03E-04	-0.095(0.014)	1.12E-11	0.011
Oily fish	1-arachidonoylglycerophosphoethanolamine*	Lysolipid	Lipid	-0.106(0.018)	1.24E-08	-0.081(0.026)	1.97E-03	-0.097(0.015)	9.46E-11	0.013
Oily fish	docosapentaenoate (n3 DPA; 22:5n3)	Essential fatty acid	Lipid	0.111(0.017)	3.62E-11	0.038(0.019)	4.72E-02	0.080(0.013)	2.22E-10	0.014
Oily fish	1-linoleoylglycerophosphoethanolamine*	Lysolipid	Lipid	-0.106(0.018)	5.45E-09	-0.055(0.025)	3.12E-02	-0.089(0.015)	1.60E-09	0.009
Oily fish	X-12627			0.093(0.016)	1.38E-08	0.044(0.019)	2.34E-02	0.073(0.012)	5.74E-09	0.009
Other fish/seafood	3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF)	Fatty acid, dicarboxylate	Lipid	0.201(0.017)	1.16E-29	0.150(0.029)	8.90E-07	0.188(0.015)	6.64E-36	0.061
Other fish/seafood	docosahexaenoate (DHA; 22:6n3)	Essential fatty acid	Lipid	0.198(0.017)	5.83E-29	0.141(0.028)	9.10E-07	0.182(0.015)	6.10E-35	0.052
Other fish/seafood	X-11469			0.189(0.017)	5.51E-27	0.131(0.031)	3.07E-05	0.175(0.015)	2.99E-31	0.052
Other fish/seafood	X-02269			0.184(0.017)	7.33E-26	0.141(0.031)	1.21E-05	0.174(0.015)	9.25E-31	0.050
Other fish/seafood	eicosapentaenoate (EPA; 20:5n3)	Essential fatty acid	Lipid	0.156(0.017)	1.71E-19	0.148(0.033)	1.38E-05	0.154(0.015)	2.66E-24	0.036
Other fish/seafood	1-docosahexaenoylglycerophosphocholine*	Lysolipid	Lipid	0.125(0.015)	7.30E-16	0.071(0.029)	1.35E-02	0.113(0.014)	6.70E-17	0.020

Table 5. List of metabolites associated with food intake from the Metabolon platform

Variable	Metabolite name	Pathway	Super-pathway	Discovery		Discordant		Meta-analysis		Whole sample
				Beta(SE)	*P*	Beta(SE)	*P*	Beta(SE)	*P*	R^2
Other fish/seafood	pyroglutamine*	Glutamate metabolism	Amino acid	-0.097(0.018)	3.81E-08	-0.102(0.025)	5.51E-05	-0.099(0.014)	5.22E-12	0.015
Other fish/seafood	docosapentaenoate (n3 DPA; 22:5n3)	Essential fatty acid	Lipid	0.086(0.015)	1.27E-08	0.102(0.030)	8.79E-04	0.089(0.013)	3.44E-11	0.010
Other fish/seafood	1-eicosatrienoylglycerophosphocholine*	Lysolipid	Lipid	-0.082(0.016)	2.73E-07	-0.083(0.032)	1.14E-02	-0.082(0.014)	8.44E-09	0.009
Other fish/seafood	1,5-anhydroglucitol (1,5-AG)	Glycolysis, gluconeogenesis, pyruvate metabolism	Carbo-hydrate	-0.079(0.016)	6.12E-07	-0.087(0.034)	1.06E-02	-0.081(0.014)	1.80E-08	0.011
Other fish/seafood	X-12696--3,4-dihydroxyphenylacetate sulfate	Phenylalanine and Tyrosine Metabolism	Amino acid	-0.084(0.017)	6.20E-07	-0.078(0.036)	2.96E-02	-0.083(0.015)	4.73E-08	0.011
Peaches	X-11315			0.158(0.024)	8.45E-11	0.097(0.024)	6.51E-05	0.127(0.017)	5.92E-14	0.039
Porridge	X-09789			0.099(0.009)	9.10E-26	0.083(0.015)	1.01E-07	0.094(0.008)	4.96E-33	0.048
Porridge	X-12253--2-aminophenol sulfate	Chemical	Xenobiotics	0.053(0.010)	3.85E-08	0.043(0.019)	2.50E-02	0.051(0.009)	2.63E-09	0.016
Poultry	pyroglutamine*	Glutamate metabolism	Amino acid	-0.102(0.016)	1.87E-10	-0.070(0.022)	1.29E-03	-0.091(0.013)	1.36E-12	0.017
Poultry	creatine	Creatine metabolism	Amino acid	0.104(0.015)	2.57E-11	0.044(0.022)	4.56E-02	0.083(0.013)	3.17E-11	0.015
Processed meats	trans-4-hydroxyproline	Urea cycle; arginine-, proline-, metabolism	Amino acid	0.052(0.009)	4.07E-09	0.035(0.016)	3.14E-02	0.048(0.008)	4.69E-10	0.011
Processed meats	X-11381			0.052(0.009)	6.89E-09	0.039(0.017)	2.18E-02	0.049(0.008)	4.74E-10	0.011
Savoury pies	X-11315			-0.287(0.044)	5.59E-11	-0.116(0.054)	3.44E-02	-0.221(0.034)	9.08E-11	0.016
Savoury pies	3-phenylpropionate (hydrocinnamate)	Phenylalanine & tyrosine metabolism	Amino acid	-0.264(0.044)	2.06E-09	-0.151(0.068)	2.89E-02	-0.231(0.037)	3.72E-10	0.013
Savoury pies	X-11372			0.290(0.047)	1.14E-09	0.114(0.055)	4.02E-02	0.216(0.036)	2.04E-09	0.015
Savoury pies	ergothioneine	Food component, Plant	Xenobiotics	-0.320(0.060)	1.31E-07	-0.291(0.105)	6.63E-03	-0.313(0.052)	2.16E-09	0.017
Savoury snacks	X-11372			0.064(0.008)	8.31E-14	0.029(0.011)	9.73E-03	0.051(0.007)	3.88E-14	0.026
Soy foods	X-11381			-0.184(0.033)	1.95E-08	-0.063(0.025)	1.46E-02	-0.108(0.020)	5.80E-08	0.018
Soy/other milk	4-ethylphenylsulfate	Benzoate metabolism	Xenobiotics	0.299(0.051)	6.10E-09	0.195(0.043)	3.18E-05	0.239(0.033)	6.05E-13	0.061
Spirits/liquor	4-androsten-3beta,17beta-diol	Sterol, Steroid	Lipid	0.044(0.008)	2.53E-08	0.028(0.013)	4.24E-02	0.040(0.007)	4.30E-09	0.021

Table 5. List of metabolites associated with food intake from the Metabolon platform

Variable	Metabolite name	Pathway	Super-pathway	Discovery		Discordant		Meta-analysis		Whole sample
				Beta(SE)	*P*	Beta(SE)	*P*	Beta(SE)	*P*	R^2
	disulfate 1*									
Sweet baked products	eicosapentaenoate (EPA; 20:5n3)	Essential fatty acid	Lipid	-0.016(0.002)	1.20E-10	-0.012(0.004)	2.30E-03	-0.014(0.002)	1.08E-12	0.016
Sweet baked products	docosahexaenoate (DHA; 22:6n3)	Essential fatty acid	Lipid	-0.016(0.002)	3.75E-11	-0.011(0.004)	7.33E-03	-0.015(0.002)	1.28E-12	0.016
Sweet baked products	scyllo-inositol	Inositol metabolism	Lipid	-0.018(0.003)	8.43E-09	-0.014(0.005)	5.30E-03	-0.017(0.003)	1.44E-10	0.019
Sweet baked products	X-11799			-0.018(0.003)	1.39E-07	-0.015(0.006)	1.37E-02	-0.017(0.003)	4.88E-09	0.020
Sweet baked products	X-11315			-0.014(0.003)	3.66E-07	-0.009(0.003)	1.26E-02	-0.012(0.002)	2.90E-08	0.014
Tomatoes	glycerate	Glycolysis, gluconeogenesis, pyruvate metabolism	Carbo-hydrate	0.029(0.006)	7.50E-07	0.018(0.007)	1.02E-02	0.025(0.005)	4.72E-08	0.009
White/brown bread, refined grains	X-11315			-0.017(0.003)	1.00E-08	-0.019(0.005)	6.53E-05	-0.017(0.002)	1.84E-12	0.017
Wholemeal bread/grains	X-09789			0.017(0.003)	3.91E-07	0.011(0.005)	3.58E-02	0.015(0.003)	6.03E-08	0.010
Wine	scyllo-inositol	Inositol metabolism	Lipid	0.056(0.004)	3.04E-43	0.036(0.007)	3.42E-06	0.052(0.003)	1.47E-49	0.117
Wine	alpha-hydroxyisovalerate	Valine, leucine and isoleucine metabolism	Amino acid	0.047(0.004)	2.09E-36	0.038(0.006)	2.32E-08	0.045(0.003)	1.23E-45	0.072
Wine	X-11795			0.042(0.003)	5.04E-34	0.030(0.007)	9.51E-05	0.039(0.003)	1.74E-38	0.062
Wine	4-androsten-3beta,17beta-diol disulfate 1*	Sterol, Steroid	Lipid	0.043(0.004)	1.09E-26	0.034(0.008)	3.31E-05	0.041(0.003)	1.50E-31	0.071
Wine	X-11799			0.035(0.004)	3.49E-22	0.028(0.006)	8.74E-06	0.033(0.003)	1.66E-27	0.052
Wine	X-10395			0.035(0.003)	9.31E-25	0.018(0.006)	3.51E-03	0.031(0.003)	3.45E-26	0.034
Wine	X-04495			0.032(0.003)	4.68E-20	0.028(0.006)	8.69E-06	0.031(0.003)	3.26E-25	0.035
Wine	5alpha-androstan-3beta,17beta-diol disulfate	Sterol, Steroid	Lipid	0.038(0.004)	6.27E-21	0.030(0.008)	4.21E-04	0.036(0.004)	2.71E-24	0.059
Wine	pipecolate	Lysine metabolism	Amino acid	0.031(0.003)	1.57E-19	0.024(0.007)	4.94E-04	0.029(0.003)	1.51E-22	0.029
Wine	eicosapentaenoate (EPA; 20:5n3)	Essential fatty acid	Lipid	0.030(0.003)	1.66E-20	0.017(0.006)	4.40E-03	0.027(0.003)	6.63E-22	0.027
Wine	docosahexaenoate (DHA; 22:6n3)	Essential fatty acid	Lipid	0.029(0.003)	2.42E-17	0.014(0.005)	1.25E-02	0.025(0.003)	6.92E-18	0.022
Wine	myo-inositol	Inositol metabolism	Lipid	0.026(0.003)	7.58E-16	0.011(0.005)	4.93E-02	0.022(0.003)	1.36E-15	0.022

Table 5. List of metabolites associated with food intake from the Metabolon platform

Variable	Metabolite name	Pathway	Super-pathway	Discovery		Discordant		Meta-analysis		Whole sample
				Beta(SE)	*P*	Beta(SE)	*P*	Beta(SE)	*P*	R^2
Wine	docosapentaenoate (n3 DPA; 22:5n3)	Essential fatty acid	Lipid	0.025(0.003)	5.38E-13	0.019(0.006)	3.46E-03	0.024(0.003)	5.28E-15	0.019
Wine	X-12627			0.024(0.003)	4.95E-14	0.014(0.006)	1.55E-02	0.022(0.003)	5.88E-15	0.017
Wine	X-01911			0.029(0.004)	7.11E-11	0.022(0.006)	1.27E-04	0.026(0.003)	2.55E-14	0.031
Wine	2-aminobutyrate	Butanoate metabolism	Amino acid	0.027(0.004)	2.83E-13	0.016(0.007)	2.04E-02	0.025(0.003)	2.73E-14	0.023
Wine	X-05907			0.024(0.003)	4.22E-12	0.017(0.005)	1.85E-03	0.022(0.003)	3.03E-14	0.019
Wine	X-10429			0.025(0.003)	1.88E-12	0.019(0.007)	7.10E-03	0.024(0.003)	3.62E-14	0.019
Wine	caprate (10:0)	Medium chain fatty acid	Lipid	0.021(0.003)	1.36E-11	0.023(0.007)	2.09E-03	0.022(0.003)	6.60E-14	0.016
Wine	piperine	Food component, Plant	Xenobiotics	0.028(0.004)	1.04E-10	0.022(0.006)	2.34E-04	0.026(0.003)	6.81E-14	0.026
Wine	X-12038			0.020(0.003)	2.27E-09	0.028(0.007)	1.20E-04	0.022(0.003)	9.20E-13	0.014
Wine	X-13215			0.023(0.004)	1.48E-09	0.023(0.006)	3.37E-04	0.023(0.003)	1.07E-12	0.016
Wine	X-11317			0.019(0.003)	3.11E-09	0.026(0.007)	1.84E-04	0.020(0.003)	1.88E-12	0.012
Wine	4-methyl-2-oxopentanoate	Valine, leucine and isoleucine metabolism	Amino acid	0.021(0.003)	1.36E-09	0.021(0.006)	1.07E-03	0.021(0.003)	3.44E-12	0.011
Wine	beta-hydroxyisovalerate	Valine, leucine and isoleucine metabolism	Amino acid	0.021(0.003)	9.19E-11	0.016(0.007)	1.64E-02	0.020(0.003)	4.44E-12	0.009
Wine	X-09026			0.020(0.003)	1.93E-09	0.019(0.006)	2.71E-03	0.020(0.003)	1.30E-11	0.011
Wine	caprylate (8:0)	Medium chain fatty acid	Lipid	0.019(0.003)	2.10E-09	0.020(0.007)	4.05E-03	0.019(0.003)	2.13E-11	0.012
Wine	benzoate	Benzoate metabolism	Xenobiotics	0.019(0.003)	4.37E-09	0.018(0.006)	2.84E-03	0.019(0.003)	3.20E-11	0.009
Wine	X-11550			0.018(0.003)	1.54E-08	0.020(0.006)	9.97E-04	0.018(0.003)	3.94E-11	0.011
Wine	stearidonate (18:4n3)	Long chain fatty acid	Lipid	0.019(0.003)	1.47E-08	0.019(0.007)	8.49E-03	0.019(0.003)	3.17E-10	0.012
Wine	2-hydroxybutyrate (AHB)	Cysteine, methionine, SAM, taurine metabolism	Amino acid	0.019(0.004)	4.61E-08	0.021(0.007)	3.45E-03	0.020(0.003)	4.15E-10	0.015
Wine	3-(4-hydroxyphenyl)lactate	Phenylalanine & tyrosine metabolism	Amino acid	0.017(0.003)	3.11E-08	0.018(0.007)	1.37E-02	0.018(0.003)	1.08E-09	0.011
Wine	3-methyl-2-oxobutyrate	Valine, leucine and isoleucine metabolism	Amino acid	0.018(0.003)	2.72E-07	0.020(0.007)	3.55E-03	0.018(0.003)	2.61E-09	0.007

Table 5. List of metabolites associated with food intake from the Metabolon platform

				Discovery		Discordant		Meta-analysis		Whole sample
Variable	Metabolite name	Pathway	Super-pathway	Beta(SE)	*P*	Beta(SE)	*P*	Beta(SE)	*P*	R^2
Wine	X-11452			0.023(0.004)	1.18E-07	0.017(0.006)	6.54E-03	0.021(0.003)	2.79E-09	0.015
Wine	epiandrosterone sulfate	Sterol, Steroid	Lipid	0.018(0.003)	2.21E-07	0.016(0.006)	1.05E-02	0.018(0.003)	6.29E-09	0.010
Wine	X-13496			0.017(0.003)	1.08E-07	0.015(0.006)	2.23E-02	0.017(0.003)	6.56E-09	0.009
Wine	arachidonate (20:4n6)	Long chain fatty acid	Lipid	0.017(0.004)	1.13E-06	0.018(0.006)	5.04E-03	0.017(0.003)	1.50E-08	0.009
Wine	theophylline	Xanthine metabolism	Xenobiotics	0.020(0.004)	2.47E-07	0.014(0.006)	2.41E-02	0.018(0.003)	1.99E-08	0.009
Wine	X-12644--1-docosahexaenoylglycerophosphoethanolamine	Lysolipid	Lipid	0.015(0.003)	8.33E-07	0.017(0.007)	2.16E-02	0.015(0.003)	4.82E-08	0.007

Notes: Table shows results of the linear regression analysis for the discovery population (excluding monozygotic twins discordant for each food group intake), the MZ discordant twin sample and the fixed effects meta-analysis of both groups. Only significant associations are shown which includes those associations passing the bonferroni cut-off in the discovery and fixed effects analyses ($1.08x10^{-6}$ = 0.05/[77 food variables x 601 detected metabolites]) and passing the 5% level of significance in the discordant twin group. The R-squared was calculated for the relevant metabolite and food group in the study population. The Metabolon platform is a non-targeted platform which identified 456 metabolites in blood; data were available for 3559 twins.

Table 6. List of metabolites associated with food intake from the Biocrates platform

Variable	Biocrates name	Pathway	Super-pathway	Discovery		Discordant		Meta-analysis		Whole sample
				Beta(SE)	*P*	Beta(SE)	*P*	Beta(SE)	*P*	R^2
Cream	lysoPhosphatidylcholine acyl C17:0	Glycerophsopholipids	Lipid	0.029(0.005)	1.36E-07	0.057(0.020)	7.68E-03	0.030(0.005)	3.36E-09	0.013
Cream	Hydroxysphingomyeline C14:1	Sphingolipids	Lipid	0.034(0.005)	8.09E-10	0.039(0.016)	2.37E-02	0.035(0.005)	2.22E-11	0.013
Cream	lysoPhosphatidylcholine acyl C28:1	Glycerophsopholipids	Lipid	0.025(0.005)	7.57E-07	0.050(0.025)	5.63E-02	0.026(0.005)	1.17E-07	0.012
Herbal tea	Phosphatidylcholine acyl-alkyl C38:3	Glycerophsopholipids	Lipid	-0.012(0.002)	1.78E-08	-0.001(0.005)	7.75E-01	-0.011(0.002)	9.91E-08	0.017
Herbal tea	Phosphatidylcholine acyl-alkyl C40:4	Glycerophsopholipids	Lipid	-0.010(0.002)	6.94E-07	-0.003(0.006)	5.44E-01	-0.009(0.002)	8.01E-07	0.014
Herbal tea	Phosphatidylcholine acyl-alkyl C40:3	Glycerophsopholipids	Lipid	-0.014(0.002)	3.64E-09	-0.004(0.006)	5.40E-01	-0.013(0.002)	5.57E-09	0.013
Herbal tea	Phosphatidylcholine acyl-alkyl C38:2	Glycerophsopholipids	Lipid	-0.018(0.003)	1.61E-07	-0.002(0.008)	8.05E-01	-0.016(0.003)	6.68E-07	0.012
Herbal tea	Phosphatidylcholine diacyl C42:4	Glycerophsopholipids	Lipid	-0.014(0.003)	1.51E-07	-0.002(0.008)	7.43E-01	-0.013(0.002)	2.80E-07	0.011
Herbal tea	Phosphatidylcholine acyl-alkyl C38:1	Glycerophsopholipids	Lipid	-0.020(0.003)	1.65E-08	-0.001(0.010)	9.33E-01	-0.018(0.003)	6.19E-08	0.010
Oily fish	Phosphatidylcholine diacyl C40:6	Glycerophsopholipids	Lipid	0.073(0.011)	1.07E-11	0.015(0.021)	4.88E-01	0.061(0.009)	7.34E-11	0.041
Oily fish	Phosphatidylcholine diacyl C38:6	Glycerophsopholipids	Lipid	0.077(0.011)	9.80E-13	0.012(0.021)	5.68E-01	0.064(0.009)	1.04E-11	0.041
White/brown bread, refined grains	Octenoylcarnitine	Acylcarnitines	Lipid	0.013(0.002)	1.32E-08	0.008(0.005)	8.23E-02	0.012(0.002)	2.45E-09	0.031
Wine	Phosphatidylcholine diacyl C36:5	Glycerophsopholipids	Lipid	0.018(0.003)	1.16E-08	0.011(0.008)	1.73E-01	0.017(0.003)	3.63E-09	0.038
Wine	Phosphatidylcholine diacyl C32:1	Glycerophsopholipids	Lipid	0.015(0.003)	2.83E-07	0.013(0.011)	2.25E-01	0.015(0.003)	9.27E-08	0.028

Notes: Table shows results of the linear regression analysis for the discovery population (excluding monozygotic twins discordant for each food group intake), the MZ discordant twin sample and the fixed effects meta-analysis of both groups. Only significant associations are shown which includes those associations passing the bonferroni cut-off in the discovery and fixed effects analyses ($1.08x10^{-6}$ = 0.05/[77 food variables x 601 detected metabolites]) and in the same direction in the discordant twin group. The R-squared was calculated for the relevant metabolite and food group in the study population. The Biocrates platform is a targeted platform which measured 163 metabolites in blood; data were available for 858 twins.

Table 7. Information for metabolites associated to food intakes, including biological role, previous dietary and disease associations

Pathway	Subpathway	Metabolite	Replicated	Novel	Biological role	Previous dietary associations	Previous disease/metabolic associations
Lip	Essential fatty acid	Docosahexaenoate (DHA; 22:6n3)	↑ Oily fish ↑ Other seafood	↑ Wine ↓ Baked goods	Essential fatty acid, derived endogenously from α-linolenic acid and in high concentrations in oily fish.	↑ Fish (excluding shellfish); ↑ Other vegetables: Celery, green beans, squash, cucumbers; ↑ White rice; ↓ Chips (Guertin et al., 2014) ↑ Fish and seafood (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014)	↓ NASH (Kalhan et al., 2011) ↓ Simvastatin treatment (Chen et al., 2011)
Lip	Essential fatty acid	Docosapentaenoate (n3 DPA; 22:5n3)	↑ Wine	↑ Other seafood ↑ Oily fish	Essential fatty acid, derived endogenously from omega-3 and 6 fatty acids and in high concentrations in oily fish.	↑ Eggs (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014) ↑ Alcohol (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014)	
Lip	Essential fatty acid	Eicosapentaenoate (EPA; 20:5n3)	↑ Oily fish ↑ Other seafood	↑ Wine ↓ Baked goods ↑ Avocado ↑ High fat salad dressings	Essential fatty acid, derived endogenously from α-linolenic acid and in high concentrations in oily fish.	↑ Fish and seafood (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014)	↓ NASH (Kalhan et al., 2011)
Lip	Lysolipid	1-Docosahexaenoylglycerophosphocholine*	↑ Oily fish ↑ Other seafood	↑ Avocado ↑ Green leafy vegetables	Downstream product of omega-3 essential fatty acid metabolism.	↑ Fish (excluding shellfish), ↑ Healthy eating index (Guertin et al., 2014)	
Lip	Lysolipid	1-docosahexaenoylglycerophosphoethanolamine		↑ Wine	Downstream product of omega-3 essential fatty acid metabolism.		
Lip	Lysolipid	1-Arachidonoylglycerophosphoethanolamine*		↓ Oily fish	Downstream product of omega-6 essential fatty acid metabolism. Lysolipids form cellular lipid bilayer, when cleaved by lipoprotein-associated phospholipase A2 form free lysophosphatidylcholines involved in inflammatory processes and may influence artherosclerotic plaque inflammation (Goncalves et al., 2012).	↓ HEI (Guertin et al., 2014)	
Lip	Lysolipid	1-Eicosatrienoylglycerophosphocholine*		↓ Oily fish ↓ Other seafood			↓ Dilated cardiomyopathy (Alexander, Lombardi, Rodriguez, Mitchell, & Marian, 2011)

Table 7. Information for metabolites associated to food intakes, including biological role, previous dietary and disease associations

Pathway	Subpathway	Metabolite	Replicated	Novel	Biological role	Previous dietary associations	Previous disease/metabolic associations
Lip	Lysolipid	1-Linoleoylglycerophosphoethanolamine*		↓ Oily fish		↓ HEI (Guertin et al., 2014)	↓ Dilated cardiomyopathy (Alexander et al., 2011)
Lip	Lysolipid	1-Oleoylglycerophosphoethanolamine		↓ Oily fish			↓ Dilated cardiomyopathy (Alexander et al., 2011)
Lip	Fatty acid, dicarboxylate	3-Carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF)	↑ Oily fish ↑ Other seafood	↑ Avocado ↑ Green leafy vegetables	Furan fatty acid, found in significant concentrations in fish. Thought to be a uremic toxin (Miyamoto et al., 2012). Though to induce beta cell dysfunction (Prentice et al., 2014).	↑ Fish and shellfish; ↑ Alliums (garlic and onions); ↑ Greens: lettuce, spinach, green pepper (Guertin et al., 2014) ↑ Fish and seafood (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014) ↓ Dietary choline depletion (Sha et al., 2010)	
Lip	Long chain fatty acid	Stearidonate (18:4n3)	↑ Wine		Derived endogenously from essential fatty acid α-linolenic acid.	↑ Alcohol (Zheng, Yu, Alexander, Steffen, Nettleton, et al., 2014)	
Lip	Long chain fatty acid	10-Nonadecenoate (19:1n9)		↑ Butter	Monounsaturate of nonadecenoate (19:0).	↓ Sugar-rich foods and beverages, dietary sucrose and carbohydrate (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014)	
Lip	Long chain fatty acid	Myristate (14:0)		↑ Butter	Contained in most animal and vegetable fats. Higher concentrations in plasma associated with heart failure (Yamagishi, Nettleton, & Folsom, 2008).	↑ Alcohol (Zheng, Yu, Alexander, Steffen, Nettleton, et al., 2014)	↓ T2D, IFG (Menni et al., 2013)
Lip	Long chain fatty acid	Pentadecanoate (15:0)	↑ Butter		Produced by ruminal bacteria in livestock, found in highest concentrations in dairy products, butter and meats from ruminant animals.	↑ Butter (Guertin et al., 2014)	↓ T2D, IFG (Menni et al., 2013)
Lip	Medium chain fatty acid	10-Undecenoate (11:1n1)	↑ Butter		Monounsaturated fatty acid found in oils and animal fats and produced endogenously.	↑ Butter (57)	↓ NASH, steatosis (51)
Lip	Fatty acid, branched	15-Methylpalmitate (isobar with 2-methylpalmitate)	↑ Butter		Methylated long chain fatty acid, found to be anti-inflammatory and anti-fibrotic in liver disease animal models (Mantawy, Tadros, Awad, Hassan, & El-Demerdash, 2012).	↑ Butter, ↓ HEI (Guertin et al., 2014)	↓ T2D, IFG (Menni et al., 2013)

Table 7. Information for metabolites associated to food intakes, including biological role, previous dietary and disease associations

Pathway	Subpathway	Metabolite	Replicated	Novel	Biological role	Previous dietary associations	Previous disease/metabolic associations
Lip	Medium chain fatty acid	Caprate (10:0)		↑ Wine	Saturated fatty acid found in oils and animal fats and produced endogenously.		
Lip	Medium chain fatty acid	Caprylate (8:0)		↑ Wine	Saturated fatty acid found in oils and animal fats and produced endogenously.		
Lip	Phosphatidyl-choline	lysoPhosphatidylcholine acyls C17:0 & C28:1		↑ Cream			
Lip	Phosphatidyl-choline	Phosphatidylcholine acyl-alkyls C38:1, C38:2, C38:3, C40:3 & C40:4		↓ Herbal tea			
Lip	Phosphatidyl-choline	Phosphatidylcholine diacyls C32:1 & C36:5		↑ Wine		↑ Factor 1: ↑ Fish and poultry, ↓ Confectionary, cake, cookies, desserts, margarine, tea, whole grain bread, pasta, rice and high fat cheese; ↑ Factor 2: ↑ sauce and butter, ↓ fish, whole grain bread, tea, grain flakes, muesli (Floegel et al., 2013)	
Lip	Phosphatidyl-choline	Phosphatidylcholine diacyls C38:6 & C40:6		↑ Oily fish		↑ Factor 1: ↑ Fish and poultry, ↓ Confectionary, cake, cookies, desserts, margarine, tea, whole grain bread, pasta, rice and high fat cheese; ↓ Factor 2: ↑ sauce and butter, ↓ fish, whole grain bread, tea, grain flakes, muesli (Floegel et al., 2013)	
Lip	Phosphatidyl-choline	Phosphatidylcholine diacyl C42:4		↓ Herbal tea			
Lip	Sphingomyelin	Hydroxysphingomyeline C14:1		↑ Cream		↑ Factor 1: ↑ Butter, garlic and coffee, ↓ Margarine, fresh fruit and soup; ↓ Factor 2: ↓ Butter, sweet bread spreads, high fat cheese, fresh fruit, whole grain bread, desserts, cake, cookies, high-fat dairy products (Floegel et al., 2013)	

Table 7. Information for metabolites associated to food intakes, including biological role, previous dietary and disease associations

Pathway	Subpathway	Metabolite	Replicated	Novel	Biological role	Previous dietary associations	Previous disease/metabolic associations
Lip	Carnitine metabolism	Nonanoylcarnitine*		↑ Butter	Ester of carnitine with pelargonic acid (C9).		
Lip	Carnitine metabolism	Octenoylcarnitine		↑ White and brown bread, refined grains	Acylcarnitines consist of carnitine and fatty acid derivatives; formed from mitochondrial beta-oxidation.	↓ Factor 1: ↑ Butter, ↓ Margarine and low fat cheese; ↑ Factor 2: ↑ Cornflakes and crisps, ↓ Fish, other vegetable fat, whole grain bread, cooked vegetables, garlic, nuts, tea, cabbage, sweet bread spreads, cake, cookies, high fat cheese (Floegel et al., 2013);	
Lip	Sterol, Steroid	4-Androsten-3beta,17beta-diol disulfate 1*	↑ Wine ↑ Spirits and liquors		Sex steroid. Found to be a testosterone precursor and androgen receptor agonist (Chen et al., 2004).	↑ Alcohol (Zheng, Yu, Alexander, Steffen, Nettleton, et al., 2014) ↓ Nuts and peanut butter, Sugar-rich foods and beverages, dietary sucrose and carbohydrate (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014) ↑ Total alcohol (Guertin et al., 2014)	
Lip	Sterol, Steroid	5-Alpha-androstan-3beta,17beta-diol disulfate	↑ Wine		Sex steroid. Arrests growth of prostate cells in mice (Weihua, Lathe, Warner, & Gustafsson, 2002), promotes growth of breast cancer cells *in vitro* (Couture, Theriault, Simard, & Labrie, 1993).	↑ Alcohol (Zheng, Yu, Alexander, Steffen, Nettleton, et al., 2014) ↓ Nuts and peanut butter, Sugar-rich foods and beverages, dietary sucrose and carbohydrate (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014) ↑ Total alcohol (Guertin et al., 2014)	
Lip	Sterol, Steroid	Epiandrosterone sulfate		↑ Wine	Sex steroid, also known as 3beta-hydroxy-5alpha-androstan-17-one sulfate. Downstream metabolite of dehydroepiandrosterone (DHEA).		
Lip	Inositol metabolism	Myo-inositol		↑ Wine	Contained in citrus fruits (Katz et al., 2011), beans, grains and nuts (Clements & Darnell, 1980).	↓ Dietary choline depletion (Sha et al., 2010)	↑ Insulin sensitivity (Gall et al., 2010) ↑ Dilated cardiomyopathy (Alexander et al., 2011)

Table 7. Information for metabolites associated to food intakes, including biological role, previous dietary and disease associations

Pathway	Subpathway	Metabolite	Replicated	Novel	Biological role	Previous dietary associations	Previous disease/metabolic associations
Lip	Inositol metabolism	Scyllo-inositol	↑ Wine	↓ Baked goods ↓ Fried fish	Contained in citrus juices and grapes; has been used as a potential Alzheimer's treatment (Ma, Thomason, & McLaurin, 2012).	↑ Citrus fruit: oranges, orange juice, grapefruit; ↑ Wine (Guertin et al., 2014) ↑ Fruit juice (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014)	
AA	Butanoate metabolism	2-Aminobutyrate	↑ Wine		Part of AHB-related pathway. AHB is a by-product of hepatic glutathione synthesis under oxidative stress (Lord & Bralley, 2008).	↑ Alcohol (Zheng, Yu, Alexander, Steffen, Nettleton, et al., 2014) ↓ Sugar-rich foods and beverages, dietary sucrose and carbohydrate (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014)	
AA	Cysteine, methionine, SAM, taurine metabolism	2-Hydroxybutyrate (AHB)	↑ Wine		Part of AHB-related pathway. AHB is a by-product of hepatic glutathione synthesis under oxidative stress (Lord & Bralley, 2008).	↑ Alcohol (Zheng, Yu, Alexander, Steffen, Nettleton, et al., 2014) ↓ Sugar-rich foods and beverages, dietary sucrose and carbohydrate (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014)	↑ T2D, IFG (Menni et al., 2013) ↓ Insulin sensitivity (Gall et al., 2010) ↑ BMI (Moore, 2013) ↑ Dilated cardiomyopathy (Alexander et al., 2011)
AA	Valine, leucine and isoleucine metabolism	alpha-Hydroxyisovalerate	↑ Wine		Part of AHB-related pathway. AHB is a by-product of hepatic glutathione synthesis under oxidative stress (Lord & Bralley, 2008).	↑ Alcohol (Zheng, Yu, Alexander, Steffen, Nettleton, et al., 2014) ↓ Sugar-rich foods and beverages, dietary sucrose and carbohydrate (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014)	↑ Dilated cardiomyopathy (Alexander et al., 2011) ↑ 28-day mortality in critically ill patients (Rogers et al., 2014)
AA	Valine, leucine and isoleucine metabolism	beta-Hydroxyisovalerate		↑ Wine	Part of AHB-related pathway. AHB is a by-product of hepatic glutathione synthesis under oxidative stress (Lord & Bralley, 2008).		
AA	Valine, leucine and isoleucine metabolism	3-Methyl-2-oxobutyrate		↑ Wine	Derived from branched chain amino acid, precursor to leucine and valine synthesis.		↑ T2D, IFG (Menni et al., 2013) ↑ BMI (Moore, 2013)
AA	Valine, leucine and isoleucine metabolism	4-Methyl-2-oxopentanoate		↑ Wine	Derived from branched chain amino acid, leucine, metabolism; accumulates, and may contribute to DNA damage and neurotoxicity, in maple-syrup urine disease (Mescka et al., 2014).		↑ T2D, IFG (Menni et al., 2013) ↑ BMI (Moore, 2013)

Table 7. Information for metabolites associated to food intakes, including biological role, previous dietary and disease associations

Pathway	Subpathway	Metabolite	Replicated	Novel	Biological role	Previous dietary associations	Previous disease/metabolic associations
AA	Lysine metabolism	Pipecolate		↑ Wine ↓ Confectionary & jams	Present in citrus fruits (Servillo et al., 2012). Thought to be formed primarily by gut bacterial lysine degradation (Fujita, Fujita, Kodama, Hada, & Higashino, 2003).		
AA	Phenylalanine & tyrosine metabolism	3-(4-Hydroxyphenyl)lactate		↑ Wine	The L-form is produced from tyrosine metabolism. The unusual D-form is of bacterial origin (Spaapen, Ketting, Wadman, Bruinvis, & Duran, 1987), particularly by Bifidobacteria and lactobacilli, from dietary phenolic compounds and has shown to reduce mitochondrial and neutrophil ROS production (Beloborodova et al., 2012).		↑ BMI (Moore, 2013)
AA	Phenylalanine & tyrosine metabolism	3-Phenylpropionate (hydrocinnamate)		↓ Fried fish ↓ Savoury pies ↑ Apples & pears	Formed by gut bacterial degradation of tyrosine (Smith & Macfarlane, 1996) and phenolic compounds (Anson et al., 2009; van Dorsten et al., 2012).		↑ NASH vitamin E responders at baseline (62)
AA	Phenylalanine & tyrosine metabolism	3,4-dihydroxyphenylacetate sulfate		↓ Other seafood			
AA	Tryptophan metabolism	Indolepropionate		↑ Bananas ↑ Apples & pears	Formed by gut bacterial degradation of tryptophan (Smith & Macfarlane, 1996). Acts as an antioxidant (Karbownik et al., 2001).	↓ Eggs, ↓ Red meat (Guertin et al., 2014) ↓ Dietary choline depletion (Sha et al., 2010)	↑ Insulin sensitivity (Gall et al., 2010) ↑ NASH vitamin E responders at baseline (Cheng, Joyce, Yates, Aouizerat, & Sanyal, 2012) ↓ Dilated cardiomyopathy (Alexander et al., 2011) ↓ Muscle mass indices in elderly (Lustgarten, Price, Chale, Phillips, & Fielding, 2014)
AA	Tryptophan metabolism	Tryptophan betaine		↑ Allium vegetables	Indole alkaloid found in legumes of *Erythrina* species, also called hypaphorine. Found to induce sleep in mice (Ozawa, Honda, Nakai, Kishida, & Ohsaki, 2008).	↑ Peanuts (Guertin et al., 2014) ↑ Nuts and peanut butter (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014)	

Table 7. Information for metabolites associated to food intakes, including biological role, previous dietary and disease associations

Pathway	Subpathway	Metabolite	Replicated	Novel	Biological role	Previous dietary associations	Previous disease/metabolic associations
AA	Creatine metabolism	Creatine		↑ Meat ↑ Poultry	Red meat a major source; vegetarians have lower circulating levels (Delanghe et al., 1989).	↓ Sugar-rich foods and beverages, dietary sucrose and carbohydrate (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014)	↓ Insulin sensitivity (Gall et al., 2010) ↓ Steatosis vs. NASH (Kalhan et al., 2011) ↑ Dilated cardiomyopathy (Alexander et al., 2011)
AA	Glutamate metabolism	Pyroglutamine*	↓ Poultry	↓ Other seafood ↓ Meat		↓ Poultry (chicken) (Guertin et al., 2014) ↑ Dietary choline depletion (Sha et al., 2010)	↑ Active TB (Weiner et al., 2012) ↑ Dilated cardiomyopathy (Alexander et al., 2011)
AA	Urea cycle; arginine-, proline-, metabolism	Trans-4-hydroxyproline		↑ Meat ↑ Processed meat ↑ Beef burgers	Major component of collagen; circulating levels increase following oral gelatin ingestion (Ohara, Matsumoto, Ito, Iwai, & Sato, 2007).	↑ Dietary choline depletion (Sha et al., 2010)	
Xeno	Xanthine metabolism	1-Methylxanthine	↑ Coffee			↑ Coffee (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014) (Guertin et al., 2015) ↑ β-carotene supplementation in smokers (Mondul et al., 2013)	
Xeno	Xanthine metabolism	7-Methylxanthine		↑ Chocolate		↑ Coffee (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014) ↑ Chocolate (Guertin et al., 2014)	
Xeno	Xanthine metabolism	Theobromine	↑ Chocolate		Bitter alkaloid characteristic of the cacao plant.	↑ Sugar-rich foods and beverages, dietary sucrose and carbohydrate (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014) ↑ Coffee (Guertin et al., 2015)	
Xeno	Xanthine metabolism	Theophylline		↑ Wine	Theophylline is methylxanthine metabolite found in caffeinated beverages and foods but also administered as a treatment for COPD and asthma. Alcohol has been shown to delay plasma clearance of theophylline (Thompson, 1992).		
Xeno	Benzoate	3-methyl catechol	↑ Coffee	↓ Black tea		↑ Coffee (Guertin et al.,	

Table 7. Information for metabolites associated to food intakes, including biological role, previous dietary and disease associations

Pathway	Subpathway	Metabolite	Replicated	Novel	Biological role	Previous dietary associations	Previous disease/metabolic associations
	metabolism	sulfate 1				2015)	
Xeno	Benzoate metabolism	Catechol sulfate	↑ Coffee		Chlorogenic acid derived metabolite (Lang et al., 2013).	↑ Coffee (Guertin et al., 2014) ↑ β-carotene supplementation in smokers (Mondul et al., 2013)	↑ Insulin sensitivity (Gall et al., 2010)
Xeno	Benzoate metabolism	O-methyl catechol sulfate		↑ Coffee			
Xeno	Benzoate metabolism	4-Ethylphenylsulfate		↑ Soy and other milks	Synthesized by gut bacteria, elevated in chronic kidney disease (Itoh, Ezawa, Kikuchi, Tsuruta, & Niwa, 2013); elevated levels may indicate intestinal permeability (Hsiao et al., 2013).	↑ Tofu (Guertin et al., 2014)	
Xeno	Benzoate metabolism	Benzoate		↑ Wine	Added to foods to prevent mold and bacterial growth (often as sodium benzoate in wine).		↓ BMI (Moore, 2013)
Xeno	Chemical	3-hydroxypyridine sulfate	↑ Coffee	↓ Black tea		↑ Coffee (Guertin et al., 2015)	
Xeno	Chemical	2-aminophenol sulfate		↑ Porridge	Benzoxazinoid metabolite excreted in urine following whole grain rye bread consumption (Bondia-Pons et al., 2013).		
Xeno	Food component, Plant	Quinate	↑ Coffee	↓ Black tea	Phenolic product of chlorogenic acid, acts as an antioxidant (Lang et al., 2013).	↑ Coffee (Guertin et al., 2014; Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014) ↓ Sugar-sweetened beverages: soda, fruit punch (Guertin et al., 2014)	↑ Insulin sensitivity
Xeno	Food component, Plant	Piperine	↑ Wine		Compound in black pepper with antioxidant effects (Srinivasan, 2014).	↑ Alcohol (Zheng, Yu, Alexander, Steffen, Nettleton, et al., 2014)	
Xeno	Food component, Plant	Ergothioneine		↑ Mushroom ↓ Savoury pies	Thiol compound which enters human cells via the organic cation transporter OCTN1. Is found in particularly high concentrations in specialty mushrooms, oat bran and beans and demonstrates protection against copper(II)-induced toxicity (Ey, Schomig, & Taubert, 2007).		↓ Parkinson's disease patients (Ey et al., 2007)

Table 7. Information for metabolites associated to food intakes, including biological role, previous dietary and disease associations

Pathway	Subpathway	Metabolite	Replicated	Novel	Biological role	Previous dietary associations	Previous disease/metabolic associations
Xeno	Food component, Plant	Stachydrine	↑ Fruit juice ↑ Citrus fruit		Present in citrus fruits and juices (Servillo, Giovane, Balestrieri, Cautela, & Castaldo, 2011).	↑ Citrus fruit: oranges, orange juice, grapefruit (Guertin et al., 2014) ↑ Fruit juice (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014) ↓ Dietary choline depletion (Sha et al., 2010)	↑ Dilated cardiomyopathy (Alexander et al., 2011)
Cho	Glycolysis, gluconeogenesis, pyruvate metabolism	Glycerate	↑ Citrus fruit	↓ Confectionary & jams ↑ Tomatoes	Contained in citrus fruits (Katz et al., 2011)	↑ Fruits and vegetables; ↑ Fruit juice (Zheng, Yu, Alexander, Steffen, & Boerwinkle, 2014)	↑ Insulin sensitivity (Gall et al., 2010)
Cho	Glycolysis, gluconeogenesis, pyruvate metabolism	1,5-Anhydroglucitol (1,5-AG)		↑ Other seafood	Marker of glycaemic control and of dietary origin (Yamanouchi et al., 1989).	↓ β-carotene supplementation in smokers (Mondul et al., 2013)	↓ T2D, IFG (Menni et al., 2013) ↓ BMI (Moore, 2013)
Cho	Nucleotide sugars, pentose metabolism	Threitol		↑ Apples and pears	Primary end product of D-xylose metabolism. Found in the edible fungus Armillaria mellea, jute and the pigeon pea plant FDB002261.		
Vit	Vitamin B6 metabolism	Pyridoxate		↑ High fibre breakfast cereals	Essential nutrient, coenzyme for synthesis of amino acids, neurotransmitters (serotonin, norepinephrine), sphingolipids, and aminolevulinic acid.	↑ Vitamins/supplements; ↑ Healthy eating index; ↑ Other fruits: plums, apricots, peaches, prunes, raisins, grapes, pineapple (Guertin et al., 2014) ↓ Dietary choline depletion (Sha et al., 2010)	
Pep	Dipeptide	Cyclo(leu-pro)		↑ Coffee ↓ Black tea	Cyclo(leu-pro) is a diketopiperazine which contributes to the bitter flavor of roasted coffee beans (Ginz & Engelhardt, 2000).	↓ Starchy vegetables: white potatoes, corn, and peas, ↑ Total alcohol (Guertin et al., 2014) ↑ Alcohol (Zheng, Yu, Alexander, Steffen, Nettleton, et al., 2014)	

Notes: Associations are reported which replicate previous findings from food intake and blood metabolite association studies and those which associations which we believe to be novel. Biological roles of most metabolites are defined and some previous disease associations with the appropriate metabolite in blood are included.

References for Table 7:

Alexander, D., Lombardi, R., Rodriguez, G., Mitchell, M. M., & Marian, A. J. (2011). Metabolomic distinction and insights into the pathogenesis of human primary dilated cardiomyopathy. Eur J Clin Invest, 41(5), 527-538. doi: 10.1111/j.1365-2362.2010.02441.x

Anson, N. M., Selinheimo, E., Havenaar, R., Aura, A. M., Mattila, I., Lehtinen, P., . . . Haenen, G. R. (2009). Bioprocessing of wheat bran improves in vitro bioaccessibility and colonic metabolism of phenolic compounds. J Agric Food Chem, 57(14), 6148-6155. doi: 10.1021/jf900492h

Beloborodova, N., Bairamov, I., Olenin, A., Shubina, V., Teplova, V., & Fedotcheva, N. (2012). Effect of phenolic acids of microbial origin on production of reactive oxygen species in mitochondria and neutrophils. J Biomed Sci, 19, 89. doi: 10.1186/1423-0127-19-89

Bondia-Pons, I., Barri, T., Hanhineva, K., Juntunen, K., Dragsted, L. O., Mykkanen, H., & Poutanen, K. (2013). UPLC-QTOF/MS metabolic profiling unveils urinary changes in humans after a whole grain rye versus refined wheat bread intervention. Mol Nutr Food Res, 57(3), 412-422. doi: 10.1002/mnfr.201200571

Chen, F., Knecht, K., Leu, C., Rutledge, S. J., Scafonas, A., Gambone, C., . . . Freedman, L. (2004). Partial agonist/antagonist properties of androstenedione and 4 androsten-3beta,17beta-diol. J Steroid Biochem Mol Biol, 91(4-5), 247-257. doi: 10.1016/j.jsbmb.2004.04.009

Chen, F., Maridakis, V., O'Neill, E. A., Hubbard, B. K., Strack, A., Beals, C., . . . Wong, P. (2011). The effects of simvastatin treatment on plasma lipid-related biomarkers in men with dyslipidaemia. Biomarkers, 16(4), 321-333. doi: 10.3109/1354750x.2011.561367

Cheng, J., Joyce, A., Yates, K., Aouizerat, B., & Sanyal, A. J. (2012). Metabolomic profiling to identify predictors of response to vitamin E for non-alcoholic steatohepatitis (NASH). PLoS One, 7(9), e44106. doi: 10.1371/journal.pone.0044106

Clements, R. S., Jr., & Darnell, B. (1980). Myo-inositol content of common foods: development of a high-myo-inositol diet. Am J Clin Nutr, 33(9), 1954-1967.

Couture, P., Theriault, C., Simard, J., & Labrie, F. (1993). Androgen receptor-mediated stimulation of 17 beta-hydroxysteroid dehydrogenase activity by dihydrotestosterone and medroxyprogesterone acetate in ZR-75-1 human breast cancer cells. Endocrinology, 132(1), 179-185. doi: 10.1210/endo.132.1.8380373

Delanghe, J., De Slypere, J. P., De Buyzere, M., Robbrecht, J., Wieme, R., & Vermeulen, A. (1989). Normal reference values for creatine, creatinine, and carnitine are lower in vegetarians. Clin Chem, 35(8), 1802-1803.

Ey, J., Schomig, E., & Taubert, D. (2007). Dietary sources and antioxidant effects of ergothioneine. J Agric Food Chem, 55(16), 6466-6474. doi: 10.1021/jf071328f

Floegel, A., von Ruesten, A., Drogan, D., Schulze, M. B., Prehn, C., Adamski, J., . . . Boeing, H. (2013). Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam. Eur J Clin Nutr, 67(10), 1100-1108. doi: 10.1038/ejcn.2013.147

Fujita, T., Fujita, M., Kodama, T., Hada, T., & Higashino, K. (2003). Determination of D- and L-pipecolic acid in food samples including processed foods. Ann Nutr Metab, 47(3-4), 165-169. doi: 70040

Gall, W. E., Beebe, K., Lawton, K. A., Adam, K. P., Mitchell, M. W., Nakhle, P. J., . . . Ferrannini, E. (2010). alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. PLoS One, 5(5), e10883. doi: 10.1371/journal.pone.0010883

Ginz, M., & Engelhardt, U. H. (2000). Identification of proline-based diketopiperazines in roasted coffee. J Agric Food Chem, 48(8), 3528-3532.
Goncalves, I., Edsfeldt, A., Ko, N. Y., Grufman, H., Berg, K., Bjorkbacka, H., . . . Nilsson, J. (2012). Evidence supporting a key role of Lp-PLA2-generated lysophosphatidylcholine in human atherosclerotic plaque inflammation. Arterioscler Thromb Vasc Biol, 32(6), 1505-1512. doi: 10.1161/atvbaha.112.249854

Guertin, K. A., Loftfield, E., Boca, S. M., Sampson, J. N., Moore, S. C., Xiao, Q., . . . Sinha, R. (2015). Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer. Am J Clin Nutr, 101(5), 1000-1011. doi: 10.3945/ajcn.114.096099

Guertin, K. A., Moore, S. C., Sampson, J. N., Huang, W. Y., Xiao, Q., Stolzenberg-Solomon, R. Z., . . . Cross, A. J. (2014). Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. Am J Clin Nutr, 100(1), 208-217. doi: 10.3945/ajcn.113.078758

Hsiao, E. Y., McBride, S. W., Hsien, S., Sharon, G., Hyde, E. R., McCue, T., . . . Mazmanian, S. K. (2013). Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. Cell, 155(7), 1451-1463. doi: 10.1016/j.cell.2013.11.024

Itoh, Y., Ezawa, A., Kikuchi, K., Tsuruta, Y., & Niwa, T. (2013). Correlation between Serum Levels of Protein-Bound Uremic Toxins in Hemodialysis Patients Measured by LC/MS/MS. Mass Spectrom (Tokyo), 2(Spec Iss), S0017. doi: 10.5702/massspectrometry.S0017

Kalhan, S. C., Guo, L., Edmison, J., Dasarathy, S., McCullough, A. J., Hanson, R. W., & Milburn, M. (2011). Plasma metabolomic profile in nonalcoholic fatty liver disease. Metabolism, 60(3), 404-413. doi: 10.1016/j.metabol.2010.03.006

Karbownik, M., Reiter, R. J., Garcia, J. J., Cabrera, J., Burkhardt, S., Osuna, C., & Lewinski, A. (2001). Indole-3-propionic acid, a melatonin-related molecule, protects hepatic microsomal membranes from iron-induced oxidative damage: relevance to cancer reduction. J Cell Biochem, 81(3), 507-513.

Katz, E., Boo, K. H., Kim, H. Y., Eigenheer, R. A., Phinney, B. S., Shulaev, V., . . . Blumwald, E. (2011). Label-free shotgun proteomics and metabolite analysis reveal a significant metabolic shift during citrus fruit development. J Exp Bot, 62(15), 5367-5384. doi: 10.1093/jxb/err197

Lang, R., Dieminger, N., Beusch, A., Lee, Y. M., Dunkel, A., Suess, B., . . . Hofmann, T. (2013). Bioappearance and pharmacokinetics of bioactives upon coffee consumption. Anal Bioanal Chem, 405(26), 8487-8503. doi: 10.1007/s00216-013-7288-0

Lord, R. S., & Bralley, J. A. (2008). Clinical applications of urinary organic acids. Part I: Detoxification markers. Altern Med Rev, 13(3), 205-215.

Lustgarten, M. S., Price, L. L., Chale, A., Phillips, E. M., & Fielding, R. A. (2014). Branched chain amino acids are associated with muscle mass in functionally limited older adults. J Gerontol A Biol Sci Med Sci, 69(6), 717-724. doi: 10.1093/gerona/glt152

Ma, K., Thomason, L. A., & McLaurin, J. (2012). scyllo-Inositol, preclinical, and clinical data for Alzheimer's disease. Adv Pharmacol, 64, 177-212. doi: 10.1016/b9780-12-394816-8.00006-4

Mantawy, E. M., Tadros, M. G., Awad, A. S., Hassan, D. A., & El-Demerdash, E. (2012). Insights antifibrotic mechanism of methyl palmitate: impact on nuclear factor kappa B and proinflammatory cytokines. Toxicol Appl Pharmacol, 258(1), 134-144. doi: 10.1016/j.taap.2011.10.016

Menni, C., Fauman, E., Erte, I., Perry, J. R., Kastenmuller, G., Shin, S. Y., . . . Spector, T. D. (2013). Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. Diabetes, 62(12), 4270-4276. doi: 10.2337/db13-0570

Mescka, C. P., Wayhs, C. A., Guerreiro, G., Manfredini, V., Dutra-Filho, C. S., & Vargas, C. R. (2014). Prevention of DNA damage by l-carnitine induced by metabolites accumulated in maple syrup urine disease in human peripheral leukocytes in vitro. Gene, 548(2), 294-298. doi: 10.1016/j.gene.2014.07.051

Miyamoto, Y., Iwao, Y., Mera, K., Watanabe, H., Kadowaki, D., Ishima, Y., . . . Maruyama, T. (2012). A uremic toxin, 3-carboxy-4-methyl-5-propyl-2-furanpropionate induces cell damage to proximal tubular cells via the generation of a radical intermediate. Biochem Pharmacol, 84(9), 1207-1214. doi: 10.1016/j.bcp.2012.07.033

Mondul, A. M., Sampson, J. N., Moore, S. C., Weinstein, S. J., Evans, A. M., Karoly, E. D., . . . Albanes, D. (2013). Metabolomic profile of response to supplementation with beta-carotene in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study. Am J Clin Nutr, 98(2), 488-493. doi: 10.3945/ajcn.113.062778

Moore, S. M., C. E.; Sampson, J. N.; Stolzenberg-Solomon, R. Z.; Zheng, W.; Cai, Q.; Tan, Y. T.; Chow, W.; Ji, B.; Liu, D. K.; Xiao, Q.; Boca, S.; Leitzmann, M. F.; Yang, G.; Xiang, Y. B.; Sinha, R.; Shu, X. O.; Cross, A. J. (2013). Human metabolic correlates of body mass index. Metabolomics, 10(2), 259-269.

Ohara, H., Matsumoto, H., Ito, K., Iwai, K., & Sato, K. (2007). Comparison of quantity and structures of hydroxyproline-containing peptides in human blood after oral ingestion of gelatin hydrolysates from different sources. J Agric Food Chem, 55(4), 1532-1535. doi: 10.1021/jf062834s

Ozawa, M., Honda, K., Nakai, I., Kishida, A., & Ohsaki, A. (2008). Hypaphorine, an indole alkaloid from Erythrina velutina, induced sleep on normal mice. Bioorg Med Chem Lett, 18(14), 3992-3994. doi: 10.1016/j.bmcl.2008.06.002

Prentice, K. J., Luu, L., Allister, E. M., Liu, Y., Jun, L. S., Sloop, K. W., . . . Wheeler, M. B. (2014). The furan fatty acid metabolite CMPF is elevated in diabetes and induces beta cell dysfunction. Cell Metab, 19(4), 653-666. doi: 10.1016/j.cmet.2014.03.008

Rogers, A. J., McGeachie, M., Baron, R. M., Gazourian, L., Haspel, J. A., Nakahira, K., . . . Choi, A. M. (2014). Metabolomic derangements are associated with mortality in critically ill adult patients. PLoS One, 9(1), e87538. doi: 10.1371/journal.pone.0087538

Servillo, L., Giovane, A., Balestrieri, M. L., Cautela, D., & Castaldo, D. (2011). Proline derivatives in fruits of bergamot (Citrus bergamia Risso et Poit): presence of N-methyl-L-proline and 4-hydroxy-L-prolinebetaine. J Agric Food Chem, 59(1), 274-281. doi: 10.1021/jf102833v

Servillo, L., Giovane, A., Balestrieri, M. L., Ferrari, G., Cautela, D., & Castaldo, D. (2012). Occurrence of pipecolic acid and pipecolic acid betaine (homostachydrine) in Citrus genus plants. J Agric Food Chem, 60(1), 315-321. doi: 10.1021/jf204286r

Sha, W., da Costa, K. A., Fischer, L. M., Milburn, M. V., Lawton, K. A., Berger, A., . . . Zeisel, S. H. (2010). Metabolomic profiling can predict which humans will develop liver dysfunction when deprived of dietary choline. Faseb j, 24(8), 2962-2975. doi: 10.1096/fj.09-154054

Smith, E. A., & Macfarlane, G. T. (1996). Enumeration of human colonic bacteria producing phenolic and indolic compounds: effects of pH, carbohydrate availability and retention time on dissimilatory aromatic amino acid metabolism. J Appl Bacteriol, 81(3), 288-302.
Spaapen, L. J., Ketting, D., Wadman, S. K., Bruinvis, L., & Duran, M. (1987). Urinary D-4-hydroxyphenyllactate, D-phenyllactate and D-2-hydroxyisocaproate, abnormalities of bacterial origin. J Inherit Metab Dis, 10(4), 383-390.

Srinivasan, K. (2014). Antioxidant potential of spices and their active constituents. Crit Rev Food Sci Nutr, 54(3), 352-372. doi: 10.1080/10408398.2011.585525

Thompson, P. J. (1992). Social nocturnal alcohol consumption and pharmacokinetics of theophylline. Chest, 101(1), 156-159.

van Dorsten, F. A., Peters, S., Gross, G., Gomez-Roldan, V., Klinkenberg, M., de Vos, R. C., . . . Jacobs, D. M. (2012). Gut microbial metabolism of polyphenols from black tea and red wine/grape juice is source-specific and colon-region dependent. J Agric Food Chem, 60(45), 11331-11342. doi: 10.1021/jf303165w

Weihua, Z., Lathe, R., Warner, M., & Gustafsson, J. A. (2002). An endocrine pathway in the prostate, ERbeta, AR, 5alpha-androstane-3beta,17beta-diol, and CYP7B1, regulates prostate growth. Proc Natl Acad Sci U S A, 99(21), 13589-13594. doi: 10.1073/pnas.162477299

Weiner, J., 3rd, Parida, S. K., Maertzdorf, J., Black, G. F., Repsilber, D., Telaar, A., . . . Kaufmann, S. H. (2012). Biomarkers of inflammation, immunosuppression and stress with active disease are revealed by metabolomic profiling of tuberculosis patients. PLoS One, 7(7), e40221. doi: 10.1371/journal.pone.0040221

Yamagishi, K., Nettleton, J. A., & Folsom, A. R. (2008). Plasma fatty acid composition and incident heart failure in middle-aged adults: the Atherosclerosis Risk in Communities (ARIC) Study. Am Heart J, 156(5), 965-974. doi: 10.1016/j.ahj.2008.06.017

Yamanouchi, T., Minoda, S., Yabuuchi, M., Akanuma, Y., Akanuma, H., Miyashita, H., & Akaoka, I. (1989). Plasma 1,5-anhydro-D-glucitol as new clinical marker of glycemic control in NIDDM patients. Diabetes, 38(6), 723-729.

Zheng, Y., Yu, B., Alexander, D., Steffen, L. M., & Boerwinkle, E. (2014). Human metabolome associates with dietary intake habits among African Americans in the atherosclerosis risk in communities study. Am J Epidemiol, 179(12), 1424-1433. doi: 10.1093/aje/kwu073

Zheng, Y., Yu, B., Alexander, D., Steffen, L. M., Nettleton, J. A., & Boerwinkle, E. (2014). Metabolomic patterns and alcohol consumption in African Americans in the Atherosclerosis Risk in Communities Study. Am J Clin Nutr, 99(6), 1470-1478. doi: 10.3945/ajcn.113.074070

Table 8. List of metabolites associated with dietary patterns from the Metabolon platform

Variable	Metabolite name	Pathway	Super-pathway	Discovery		Discordant		Meta-analysis		Unique to pattern
				Beta(SE)	*P*	Beta(SE)	*P*	Beta(SE)	*P*	
Mediterranean Diet Score	X-11315			0.117(0.010)	3.22E-28	0.085(0.018)	2.01E-06	0.109(0.009)	7.94E-34	No
Mediterranean Diet Score	X-11469			0.104(0.010)	3.86E-23	0.091(0.016)	3.08E-08	0.100(0.009)	9.16E-31	No
Mediterranean Diet Score	X-02269			0.101(0.010)	7.10E-22	0.093(0.016)	2.98E-08	0.099(0.009)	1.70E-29	No
Mediterranean Diet Score	docosahexaenoate (DHA; 22:6n3)	Essential fatty acid	Lipid	0.100(0.010)	5.78E-21	0.090(0.017)	3.88E-07	0.097(0.009)	2.45E-27	No
Mediterranean Diet Score	3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF)	Fatty acid, dicarboxylate	Lipid	0.084(0.010)	4.57E-17	0.082(0.016)	2.72E-07	0.084(0.008)	1.82E-23	No
Mediterranean Diet Score	eicosapentaenoate (EPA; 20:5n3)	Essential fatty acid	Lipid	0.089(0.010)	2.82E-18	0.062(0.019)	9.62E-04	0.083(0.009)	9.54E-21	No
Mediterranean Diet Score	1-docosahexaenoylglycerophosphocholine*	Lysolipid	Lipid	0.084(0.011)	2.16E-15	0.068(0.018)	2.11E-04	0.080(0.009)	1.28E-18	No
Mediterranean Diet Score	indolepropionate	Tryptophan metabolism	Amino acid	0.086(0.011)	7.59E-16	0.054(0.018)	3.44E-03	0.078(0.009)	1.60E-17	No
Mediterranean Diet Score	X-21365 [trimethyl-N-aminovalerate]			-0.076(0.011)	1.60E-12	-0.064(0.018)	3.20E-04	-0.073(0.009)	1.48E-15	No
Mediterranean Diet Score	tryptophan betaine	Tryptophan metabolism	Amino acid	0.089(0.012)	2.36E-13	0.068(0.022)	2.53E-03	0.084(0.011)	1.65E-15	No
Mediterranean Diet Score	X-11847			0.091(0.013)	9.97E-13	0.074(0.024)	2.38E-03	0.088(0.011)	6.32E-15	Yes
Mediterranean Diet Score	X-11381			-0.075(0.011)	5.75E-12	-0.052(0.018)	3.27E-03	-0.069(0.009)	8.27E-14	No
Mediterranean Diet Score	X-11849			0.094(0.014)	3.05E-11	0.067(0.023)	3.69E-03	0.086(0.012)	4.15E-13	Yes
Mediterranean Diet Score	X-12798			-0.067(0.011)	2.68E-09	-0.076(0.019)	8.03E-05	-0.069(0.010)	6.80E-13	No
Mediterranean Diet Score	ergothioneine	Food component, Plant	Xenobiotics	0.083(0.016)	1.08E-07	0.107(0.026)	4.21E-05	0.090(0.013)	1.56E-11	No
Mediterranean Diet Score	X-13477			0.060(0.012)	4.66E-07	0.086(0.021)	6.66E-05	0.067(0.010)	1.56E-10	Yes
Mediterranean Diet Score	scyllo-inositol	Inositol metabolism	Lipid	0.067(0.012)	1.52E-08	0.061(0.022)	5.19E-03	0.066(0.010)	2.25E-10	No
Mediterranean Diet Score	piperine	Food component, Plant	Xenobiotics	0.060(0.012)	4.12E-07	0.050(0.020)	1.39E-02	0.057(0.010)	1.73E-08	No
Fruit & Vegetable	X-11315			0.145(0.009)	5.48E-52	0.130(0.022)	1.20E-08	0.142(0.008)	3.87E-63	No
Fruit & Vegetable	X-11469			0.112(0.010)	3.63E-29	0.083(0.018)	1.16E-05	0.106(0.009)	3.39E-34	No
Fruit & Vegetable	X-11372			-0.109(0.009)	6.97E-32	-0.061(0.021)	4.25E-03	-0.102(0.008)	5.27E-34	No

Table 8. List of metabolites associated with dietary patterns from the Metabolon platform

Variable	Metabolite name	Pathway	Super-pathway	Discovery		Discordant		Meta-analysis		Unique to pattern
				Beta(SE)	*P*	Beta(SE)	*P*	Beta(SE)	*P*	
Fruit & Vegetable	X-02269			0.108(0.010)	1.21E-26	0.079(0.019)	5.22E-05	0.102(0.009)	7.46E-31	No
Fruit & Vegetable	glycerate	Glycolysis, gluconeogenesis, pyruvate metabolism	Carbo-hydrate	0.094(0.010)	1.57E-21	0.094(0.019)	1.78E-06	0.094(0.009)	1.97E-27	No
Fruit & Vegetable	eicosapentaenoate (EPA; 20:5n3)	Essential fatty acid	Lipid	0.105(0.011)	2.63E-22	0.069(0.018)	1.66E-04	0.095(0.009)	1.86E-25	No
Fruit & Vegetable	scyllo-inositol	Inositol metabolism	Lipid	0.093(0.010)	1.08E-20	0.068(0.021)	1.08E-03	0.088(0.009)	2.22E-23	No
Fruit & Vegetable	ergothioneine	Food component, Plant	Xenobiotics	0.117(0.013)	7.72E-20	0.080(0.030)	8.94E-03	0.111(0.012)	8.16E-22	No
Fruit & Vegetable	docosahexaenoate (DHA; 22:6n3)	Essential fatty acid	Lipid	0.102(0.012)	5.49E-18	0.067(0.018)	1.91E-04	0.091(0.010)	6.05E-21	No
Fruit & Vegetable	threonate	Ascorbate and aldarate metabolism	Cofactors and vitamins	0.080(0.009)	5.50E-17	0.090(0.022)	6.95E-05	0.081(0.009)	6.46E-21	Yes
Fruit & Vegetable	indolepropionate	Tryptophan metabolism	Amino acid	0.081(0.010)	1.81E-16	0.085(0.020)	3.00E-05	0.082(0.009)	8.42E-21	No
Fruit & Vegetable	1-docosahexaenoylglycerophosphocholine*	Lysolipid	Lipid	0.077(0.009)	8.09E-18	0.048(0.020)	2.03E-02	0.072(0.008)	5.51E-19	No
Fruit & Vegetable	3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF)	Fatty acid, dicarboxylate	Lipid	0.082(0.010)	9.19E-17	0.053(0.020)	7.55E-03	0.076(0.009)	2.74E-18	No
Fruit & Vegetable	tryptophan betaine	Tryptophan metabolism	Amino acid	0.083(0.010)	9.57E-16	0.064(0.025)	1.06E-02	0.080(0.009)	1.94E-17	No
Fruit & Vegetable	X-14056			-0.071(0.010)	2.40E-13	-0.081(0.021)	1.29E-04	-0.073(0.009)	6.78E-17	Yes
Fruit & Vegetable	X-12056			0.094(0.012)	1.19E-14	0.086(0.029)	3.95E-03	0.093(0.011)	6.91E-17	Yes
Fruit & Vegetable	1,5-anhydroglucitol (1,5-AG)	Glycolysis, gluconeogenesis, pyruvate metabolism	Carbo-hydrate	-0.074(0.009)	2.92E-15	-0.054(0.021)	1.32E-02	-0.071(0.009)	1.02E-16	No
Fruit & Vegetable	3-phenylpropionate (hydrocinnamate)	Phenylalanine & tyrosine metabolism	Amino acid	0.077(0.011)	4.84E-12	0.089(0.020)	1.72E-05	0.080(0.010)	1.73E-16	No
Fruit & Vegetable	X-11847			0.077(0.011)	2.85E-11	0.092(0.025)	3.65E-04	0.080(0.010)	2.68E-14	Yes
Fruit & Vegetable	X-11849			0.085(0.013)	1.96E-11	0.083(0.027)	2.38E-03	0.085(0.011)	9.86E-14	Yes
Fruit & Vegetable	pipecolate	Lysine metabolism	Amino acid	0.060(0.009)	6.41E-11	0.057(0.020)	3.94E-03	0.059(0.008)	6.30E-13	No
Fruit & Vegetable	hippurate	Benzoate metabolism	Xenobiotics	0.061(0.009)	2.42E-11	0.042(0.021)	4.86E-02	0.058(0.008)	3.29E-12	Yes

Table 8. List of metabolites associated with dietary patterns from the Metabolon platform

Variable	Metabolite name	Pathway	Super-pathway	Discovery		Discordant		Meta-analysis		Unique to pattern
				Beta(SE)	*P*	Beta(SE)	*P*	Beta(SE)	*P*	
Fruit & Vegetable	X-09789			0.062(0.010)	7.79E-10	0.059(0.025)	1.84E-02	0.061(0.009)	3.50E-11	No
Fruit & Vegetable	catechol sulfate	Benzoate metabolism	Xenobiotics	0.058(0.010)	1.78E-08	0.071(0.021)	7.28E-04	0.060(0.009)	4.21E-11	No
Fruit & Vegetable	myo-inositol	Inositol metabolism	Lipid	0.051(0.009)	5.14E-09	0.056(0.019)	3.08E-03	0.052(0.008)	4.43E-11	No
Fruit & Vegetable	threitol	Nucleotide sugars, pentose metabolism	Carbo-hydrate	0.056(0.010)	6.64E-09	0.058(0.021)	6.08E-03	0.056(0.009)	1.07E-10	No
Fruit & Vegetable	X-11261			-0.062(0.011)	2.39E-08	-0.057(0.019)	3.08E-03	-0.061(0.010)	2.08E-10	Yes
Fruit & Vegetable	X-12063			-0.047(0.009)	1.73E-07	-0.052(0.018)	4.69E-03	-0.048(0.008)	2.31E-09	Yes
Fruit & Vegetable	X-11858			0.080(0.015)	9.59E-08	0.077(0.034)	2.48E-02	0.080(0.014)	5.28E-09	No
Fruit & Vegetable	proline	Urea cycle; arginine-, proline-, metabolism	Amino acid	-0.053(0.010)	2.60E-07	-0.060(0.027)	2.87E-02	-0.054(0.010)	1.95E-08	Yes
Fruit & Vegetable	N-acetylornithine	Urea cycle; arginine-, proline-, metabolism	Amino acid	0.049(0.010)	3.45E-07	0.052(0.024)	2.92E-02	0.050(0.009)	2.57E-08	Yes
Fruit & Vegetable	gamma-glutamylvaline	gamma-glutamyl	Peptide	-0.047(0.009)	1.04E-06	-0.043(0.018)	1.71E-02	-0.046(0.008)	4.79E-08	Yes
High Alcohol	alpha-hydroxyisovalerate	Valine, leucine and isoleucine metabolism	Amino acid	0.165(0.014)	1.32E-32	0.108(0.036)	3.05E-03	0.158(0.013)	2.20E-35	No
High Alcohol	X-11795			0.150(0.013)	1.93E-28	0.118(0.026)	1.02E-05	0.143(0.012)	1.19E-33	No
High Alcohol	scyllo-inositol	Inositol metabolism	Lipid	0.146(0.015)	2.89E-22	0.128(0.033)	1.54E-04	0.143(0.014)	4.20E-26	No
High Alcohol	piperine	Food component, Plant	Xenobiotics	0.133(0.015)	1.91E-18	0.084(0.026)	1.63E-03	0.121(0.013)	1.62E-20	No
High Alcohol	4-androsten-3beta,17beta-diol disulfate 1*	Sterol, Steroid	Lipid	0.113(0.015)	9.34E-14	0.157(0.031)	1.35E-06	0.122(0.014)	3.27E-19	No
High Alcohol	X-09789			-0.118(0.014)	6.15E-16	-0.107(0.030)	4.23E-04	-0.116(0.013)	4.99E-19	No
High Alcohol	ergothioneine	Food component, Plant	Xenobiotics	0.147(0.019)	1.65E-14	0.147(0.037)	1.21E-04	0.147(0.017)	2.44E-18	No
High Alcohol	X-11799			0.116(0.015)	9.27E-14	0.137(0.032)	4.04E-05	0.120(0.014)	6.07E-18	No
High Alcohol	X-01911			0.129(0.016)	7.26E-16	0.074(0.027)	7.15E-03	0.115(0.014)	3.77E-17	No
High Alcohol	tryptophan betaine	Tryptophan metabolism	Amino acid	0.129(0.016)	6.70E-16	0.069(0.030)	2.27E-02	0.116(0.014)	1.08E-16	No
High Alcohol	X-12798			-0.103(0.014)	8.77E-13	-0.086(0.025)	6.77E-04	-0.099(0.012)	1.58E-15	No

Table 8. List of metabolites associated with dietary patterns from the Metabolon platform

Variable	Metabolite name	Pathway	Super-pathway	Discovery		Discordant		Meta-analysis		Unique to pattern
				Beta(SE)	*P*	Beta(SE)	*P*	Beta(SE)	*P*	
High Alcohol	5alpha-androstan-3beta,17beta-diol disulfate	Sterol, Steroid	Lipid	0.109(0.015)	1.02E-12	0.130(0.038)	7.44E-04	0.112(0.014)	1.73E-15	No
High Alcohol	X-21365 [trimethyl-N-aminovalerate]			-0.083(0.013)	1.72E-10	-0.123(0.026)	4.64E-06	-0.091(0.012)	3.91E-15	No
High Alcohol	gamma-tocopherol	Tocopherol metabolism	Cofactors and vitamins	0.107(0.015)	4.11E-13	0.062(0.029)	3.39E-02	0.098(0.013)	6.75E-14	Yes
High Alcohol	2-aminobutyrate	Butanoate metabolism	Amino acid	0.088(0.014)	1.23E-10	0.109(0.030)	2.85E-04	0.092(0.012)	1.08E-13	No
High Alcohol	X-10395			0.097(0.014)	3.99E-12	0.066(0.030)	2.90E-02	0.091(0.013)	3.80E-13	No
High Alcohol	X-04495			0.090(0.014)	4.83E-11	0.079(0.026)	3.16E-03	0.088(0.012)	4.06E-13	No
High Alcohol	X-11381			-0.073(0.013)	2.04E-08	-0.096(0.028)	6.04E-04	-0.077(0.012)	4.54E-11	No
High Alcohol	X-11847			0.087(0.015)	1.42E-08	0.075(0.035)	3.16E-02	0.085(0.014)	1.11E-09	Yes
High Alcohol	4-vinylphenol sulfate	Benzoate metabolism	Xenobiotics	0.078(0.014)	7.28E-08	0.074(0.027)	7.55E-03	0.077(0.013)	1.50E-09	Yes
High Alcohol	10-undecenoate (11:1n1)	Medium chain fatty acid	Lipid	0.072(0.014)	4.15E-07	0.075(0.028)	7.09E-03	0.073(0.013)	8.17E-09	No
High Alcohol	X-12816			0.093(0.017)	1.19E-07	0.075(0.036)	3.97E-02	0.089(0.016)	1.22E-08	No
High Alcohol	X-12038			0.065(0.013)	9.44E-07	0.057(0.028)	4.23E-02	0.064(0.012)	1.02E-07	No
Traditional English	stachydrine	Food component, Plant	Xenobiotics	-0.100(0.013)	1.17E-13	-0.076(0.025)	2.74E-03	-0.094(0.012)	9.94E-16	No
Traditional English	X-11315			-0.099(0.017)	4.61E-09	-0.082(0.025)	1.08E-03	-0.093(0.014)	1.68E-11	No
Traditional English	3-phenylpropionate (hydrocinnamate)	Phenylalanine & tyrosine metabolism	Amino acid	-0.083(0.016)	4.19E-07	-0.089(0.025)	4.17E-04	-0.084(0.014)	5.18E-10	No
Traditional English	X-11372			0.083(0.016)	2.83E-07	0.083(0.024)	7.93E-04	0.083(0.013)	6.68E-10	No
Traditional English	creatine	Creatine metabolism	Amino acid	0.089(0.015)	6.26E-09	0.050(0.022)	2.63E-02	0.076(0.013)	1.19E-09	No
Traditional English	X-11381			0.072(0.014)	1.84E-07	0.070(0.026)	7.14E-03	0.072(0.012)	3.71E-09	No
Dieting	X-14473			-0.143(0.017)	3.69E-17	-0.161(0.028)	2.78E-08	-0.148(0.014)	8.46E-25	No
Dieting	X-02249			0.097(0.016)	8.79E-10	0.072(0.029)	1.35E-02	0.091(0.014)	4.01E-11	No
Dieting	quinate	Food component, Plant	Xenobiotics	-0.098(0.017)	1.70E-08	-0.088(0.029)	2.86E-03	-0.095(0.015)	1.40E-10	No
Dieting	4-ethylphenylsulfate	Benzoate metabolism	Xenobiotics	0.104(0.019)	1.01E-07	0.093(0.039)	1.84E-02	0.102(0.017)	4.89E-09	No
Dieting	X-14374			-0.088(0.017)	1.27E-07	-0.065(0.031)	3.85E-02	-0.083(0.015)	1.48E-08	No
Low Meat	3-carboxy-4-methyl-5-	Fatty acid,	Lipid	-0.145(0.017)	4.24E-17	-0.139(0.034)	5.26E-05	-0.144(0.015)	3.63E-21	No

Table 8. List of metabolites associated with dietary patterns from the Metabolon platform

Variable	Metabolite name	Pathway	Super-pathway	Discovery		Discordant		Meta-analysis		Unique to pattern
				Beta(SE)	*P*	Beta(SE)	*P*	Beta(SE)	*P*	
	propyl-2-furanpropanoate (CMPF)	dicarboxylate								
Low Meat	pyroglutamine*	Glutamate metabolism	Amino acid	0.102(0.016)	1.62E-10	0.162(0.024)	1.28E-10	0.120(0.013)	8.99E-20	No
Low Meat	docosahexaenoate (DHA; 22:6n3)	Essential fatty acid	Lipid	-0.134(0.017)	4.71E-15	-0.128(0.030)	2.97E-05	-0.133(0.015)	2.65E-19	No
Low Meat	eicosapentaenoate (EPA; 20:5n3)	Essential fatty acid	Lipid	-0.131(0.016)	2.89E-15	-0.102(0.026)	9.59E-05	-0.123(0.014)	8.44E-19	No
Low Meat	2-aminobutyrate	Butanoate metabolism	Amino acid	-0.114(0.017)	1.57E-11	-0.108(0.028)	1.89E-04	-0.113(0.015)	7.47E-15	No
Low Meat	X-11261			0.111(0.016)	1.30E-11	0.089(0.027)	9.83E-04	0.105(0.014)	4.20E-14	Yes
Low Meat	X-11469			-0.111(0.017)	2.74E-11	-0.106(0.035)	2.73E-03	-0.110(0.015)	1.88E-13	No
Low Meat	X-02269			-0.112(0.017)	2.97E-11	-0.097(0.032)	3.24E-03	-0.109(0.015)	2.61E-13	No
Low Meat	creatine	Creatine metabolism	Amino acid	-0.092(0.015)	2.80E-09	-0.117(0.028)	4.24E-05	-0.098(0.014)	4.12E-13	No
Low Meat	X-02249			-0.122(0.018)	6.73E-12	-0.073(0.031)	1.99E-02	-0.110(0.015)	7.70E-13	No
Low Meat	X-11372			0.112(0.017)	1.24E-10	0.076(0.024)	1.73E-03	0.099(0.014)	1.17E-12	No
Low Meat	alpha-hydroxyisovalerate	Valine, leucine and isoleucine metabolism	Amino acid	-0.106(0.016)	1.24E-10	-0.078(0.027)	4.42E-03	-0.099(0.014)	2.09E-12	No
Low Meat	15-methylpalmitate (isobar with 2-methylpalmitate)	Fatty acid, branched	Lipid	-0.109(0.019)	5.20E-09	-0.091(0.029)	2.06E-03	-0.104(0.016)	3.23E-11	No
Low Meat	X-11478			0.103(0.017)	3.48E-09	0.073(0.027)	6.92E-03	0.094(0.015)	9.86E-11	Yes
Low Meat	2-hydroxybutyrate (AHB)	Cysteine, methionine, SAM, taurine metabolism	Amino acid	-0.081(0.016)	1.70E-07	-0.095(0.028)	7.53E-04	-0.085(0.014)	4.08E-10	No
Low Meat	betaine	Glycine, serine and threonine metabolism	Amino acid	0.087(0.017)	1.90E-07	0.078(0.029)	7.05E-03	0.085(0.014)	3.87E-09	Yes
Low Meat	C-glycosyltryptophan*	Tryptophan metabolism	Amino acid	0.072(0.014)	3.27E-07	0.056(0.024)	1.77E-02	0.068(0.012)	1.80E-08	Yes
Low Meat	X-11521			0.086(0.017)	2.79E-07	0.064(0.028)	2.32E-02	0.080(0.014)	2.12E-08	Yes
Low Meat	X-11204			-0.081(0.016)	6.79E-07	-0.063(0.025)	1.12E-02	-0.076(0.014)	2.52E-08	Yes

Notes: Table shows results of the linear regression analysis for the discovery population (excluding monozygotic twins discordant for each diet pattern), the MZ discordant twin sample and the fixed effects meta-analysis of both groups. Only significant associations are shown which includes those associations passing the bonferroni cut-off in the discovery and fixed effects analyses (1.08×10^{-6} = 0.05/[77 diet phenotypes x 601 detected metabolites]) and passing the 5% level of significance in the discordant twin group. The Metabolon platform is a non-targeted platform which identified 456 metabolites in blood for which data were available for 3559 twins.

Table 9. List of metabolites associated with dietary patterns from the Metabolon platform.

Variable	Metabolite	Pathway	Super-pathway	Discovery		Discordant		Meta-analysis		Unique to patterns
				Beta(SE)	*P*	Beta(SE)	*P*	Beta(SE)	*P*	
Fruit & Vegetable	Octenoylcarnitine	Acylcarnitines	Lipid	-0.052(0.009)	1.96E-08	-0.021(0.018)	0.261	-0.045(0.008)	2.09E-08	No
Low Meat	Phosphatidylcholine diacyl C36:5	Glycerophospholipids	Lipid	-0.115(0.016)	8.26E-12	-0.028(0.032)	0.380	-0.096(0.015)	3.96E-11	No
Low Meat	Phosphatidylcholine diacyl C36:6	Glycerophospholipids	Lipid	-0.094(0.016)	3.81E-09	-0.024(0.036)	0.510	-0.083(0.014)	8.53E-09	Yes
Low Meat	Phosphatidylcholine acyl-alkyl C38:6	Glycerophospholipids	Lipid	-0.061(0.011)	9.29E-08	-0.016(0.023)	0.481	-0.053(0.010)	2.33E-07	Yes
Low Meat	Phosphatidylcholine diacyl C38:6	Glycerophospholipids	Lipid	-0.067(0.013)	6.80E-07	-0.037(0.035)	0.294	-0.063(0.012)	3.91E-07	No

Notes: Table shows results of the linear regression analysis for the discovery population (excluding monozygotic twins discordant for each diet pattern), the MZ discordant twin sample and the fixed effects meta-analysis of both groups. Only significant associations are shown which includes those associations passing the bonferroni cut-off in the discovery and fixed effects analyses ($1.08x10^{-6}$ = 0.05/[77 diet phenotypes x 601 detected metabolites]) and in the same direction in the discordant twin group. The Biocrates platform is a targeted platform which measures 163 metabolites in blood for which data were available for 858 twins.

Table 10. SNP associations with diet-associated metabolites from previous genome-wide association study (Shin et al., Nat Genet. 2014 Jun;46(6):543-50)

			Diet variable-Metabolite (Meta-analysis result)						SNP-Metabolite (Shin et al., 2014)	
Metabolite	**Diet variable**	**Diet pattern or food**	**Beta(SE)**	***P***	**Gene symbol**	**SNP**	**EA/OA**	**EAF**	**Beta(SE)**	***P***
2-Aminobutyrate	Low Meat	Pattern	-0.113(0.015)	7.47E-15	*PPM1K*	rs10022462	T/C	0.45	-0.012(0.002)	4.55E-11
2-Aminobutyrate	Wine	Food	0.025(0.003)	2.73E-14	*SLC1A4*	rs10211524	A/G	0.42	0.019(0.002)	5.59E-16
2-Aminobutyrate	Wine	Food	0.025(0.003)	2.73E-14	*PPM1K*	rs10022462	T/C	0.45	-0.012(0.002)	4.55E-11
2-Aminobutyrate	High Alcohol	Pattern	0.092(0.012)	1.08E-13	*SLC1A4*	rs10211524	A/G	0.42	0.019(0.002)	5.59E-16
3-(4-Hydroxyphenyl)lactate	Wine	Food	0.018(0.003)	1.08E-09	*CCDC57*	rs4625783	T/C	0.43	-0.019(0.003)	3.90E-13
3-Phenylpropionate (hydrocinnamate)	Fruit & Vegetable	Pattern	0.080(0.010)	1.73E-16	*ACSM5*	rs11647589	A/G	0.72	0.030(0.005)	2.85E-11
3-Phenylpropionate (hydrocinnamate)	Fried fish	Food	-0.172(0.026)	4.12E-11	*ACSM5*	rs11647589	A/G	0.72	0.030(0.005)	2.85E-11
3-Phenylpropionate (hydrocinnamate)	Savoury pies	Food	-0.231(0.037)	3.72E-10	*ACSM5*	rs11647589	A/G	0.72	0.030(0.005)	2.85E-11
3-Phenylpropionate (hydrocinnamate)	Traditional English	Pattern	-0.084(0.014)	5.18E-10	*ACSM5*	rs11647589	A/G	0.72	0.030(0.005)	2.85E-11
3-Phenylpropionate (hydrocinnamate)	Apples & pears	Food	0.024(0.004)	1.24E-08	*ACSM5*	rs11647589	A/G	0.72	0.030(0.005)	2.85E-11
alpha-Hydroxyisovalerate	Wine	Food	0.045(0.003)	1.23E-45	*HAO2*	rs12141041	T/C	0.47	-0.025(0.004)	1.77E-12
alpha-Hydroxyisovalerate	High Alcohol	Pattern	0.158(0.013)	2.20E-35	*HAO2*	rs12141041	T/C	0.47	-0.025(0.004)	1.77E-12
alpha-Hydroxyisovalerate	Low Meat	Pattern	-0.099(0.014)	2.09E-12	*HAO2*	rs12141041	T/C	0.47	-0.025(0.004)	1.77E-12
Betaine	Low Meat	Pattern	0.085(0.014)	3.87E-09	*CBS*	rs2851391	T/C	0.46	-0.012(0.002)	1.15E-11
Creatine	Low Meat	Pattern	-0.098(0.014)	4.12E-13	*CPS1*	rs715	T/C	0.71	-0.045(0.004)	9.63E-25
Creatine	Meat	Food	0.063(0.009)	8.24E-12	*CPS1*	rs715	T/C	0.71	-0.045(0.004)	9.63E-25
Creatine	Poultry	Food	0.083(0.013)	3.17E-11	*CPS1*	rs715	T/C	0.71	-0.045(0.004)	9.63E-25
Creatine	Traditional English	Pattern	0.076(0.013)	1.19E-09	*CPS1*	rs715	T/C	0.71	-0.045(0.004)	9.63E-25
Indolepropionate	Fruit & Vegetable	Pattern	0.082(0.009)	8.42E-21	*ACSM2A*	rs1394678	T/C	0.28	-0.035(0.004)	1.70E-20
Indolepropionate	Mediterranean Diet Score	Pattern	0.078(0.009)	1.60E-17	*ACSM2A*	rs1394678	T/C	0.28	-0.035(0.004)	1.70E-20
Indolepropionate	Bananas	Food	0.034(0.005)	1.05E-11	*ACSM2A*	rs1394678	T/C	0.28	-0.035(0.004)	1.70E-20
Indolepropionate	Apples & pears	Food	0.026(0.004)	2.39E-09	*ACSM2A*	rs1394678	T/C	0.28	-0.035(0.004)	1.70E-20
N-acetylornithine	Fruit & Vegetable	Pattern	0.050(0.009)	2.57E-08	*NAT8*	rs10206899	T/C	0.78	0.221(0.005)	4.66E-481
proline	Fruit & Vegetable	Pattern	-0.054(0.010)	1.95E-08	*PRODH*	rs2540641	A/C	0.08	0.063(0.004)	2.98E-59

Table 10. SNP associations with diet-associated metabolites from previous genome-wide association study (Shin et al., Nat Genet. 2014 Jun;46(6):543-50)

			Diet variable-Metabolite (Meta-analysis result)						SNP-Metabolite (Shin et al., 2014)	
Metabolite	**Diet variable**	**Diet pattern or food**	**Beta(SE)**	***P***	**Gene symbol**	**SNP**	**EA/OA**	**EAF**	**Beta(SE)**	***P***
Pyroglutamine*	Low Meat	Pattern	0.120(0.013)	8.99E-20	*SLC6A13*	rs11613331	A/G	0.55	0.037(0.004)	2.23E-25
Pyroglutamine*	Low Meat	Pattern	0.120(0.013)	8.99E-20	*SLC6A20*	rs17279437	A/G	0.10	0.059(0.006)	1.25E-20
Pyroglutamine*	Poultry	Food	-0.091(0.013)	1.36E-12	*SLC6A13*	rs11613331	A/G	0.55	0.037(0.004)	2.23E-25
Pyroglutamine*	Poultry	Food	-0.091(0.013)	1.36E-12	*SLC6A20*	rs17279437	A/G	0.10	0.059(0.006)	1.25E-20
Pyroglutamine*	Other seafood	Food	-0.099(0.014)	5.22E-12	*SLC6A13*	rs11613331	A/G	0.55	0.037(0.004)	2.23E-25
Pyroglutamine*	Other seafood	Food	-0.099(0.014)	5.22E-12	*SLC6A20*	rs17279437	A/G	0.10	0.059(0.006)	1.25E-20
Pyroglutamine*	Meat	Food	-0.062(0.010)	2.10E-10	*SLC6A13*	rs11613331	A/G	0.55	0.037(0.004)	2.23E-25
Pyroglutamine*	Meat	Food	-0.062(0.010)	2.10E-10	*SLC6A20*	rs17279437	A/G	0.10	0.059(0.006)	1.25E-20
Tryptophan betaine	Fruit & Vegetable	Pattern	0.080(0.009)	1.94E-17	*SLC22A4*	rs2405522	A/G	0.17	-0.126(0.011)	4.49E-29
Tryptophan betaine	High Alcohol	Pattern	0.116(0.014)	1.08E-16	*SLC22A4*	rs2405522	A/G	0.17	-0.126(0.011)	4.49E-29
Tryptophan betaine	Mediterranean Diet Score	Pattern	0.084(0.011)	1.65E-15	*SLC22A4*	rs2405522	A/G	0.17	-0.126(0.011)	4.49E-29
Tryptophan betaine	Allium vegetables	Food	0.034(0.006)	4.73E-09	*SLC22A4*	rs2405522	A/G	0.17	-0.126(0.011)	4.49E-29
1,5-anhydroglucitol (1,5-AG)	Fruit & Vegetable	Pattern	-0.071(0.009)	1.02E-16	*RAB3GAP1*	rs7570971	A/C	0.33	-0.037(0.003)	7.86E-45
1,5-Anhydroglucitol (1,5-AG)	Fruit & Vegetable	Pattern	-0.071(0.009)	1.02E-16	*MGAM*	rs3800993	T/C	0.17	0.029(0.004)	1.53E-15
1,5-Anhydroglucitol (1,5-AG)	Other seafood	Food	-0.081(0.014)	1.80E-08	*RAB3GAP1*	rs7570971	A/C	0.33	-0.037(0.003)	7.86E-45
1,5-Anhydroglucitol (1,5-AG)	Other seafood	Food	-0.081(0.014)	1.80E-08	*MGAM*	rs3800993	T/C	0.17	0.029(0.004)	1.53E-15
1-Arachidonoylglycerophosphoethanolamine*	Oily fish	Food	-0.097(0.015)	9.46E-11	*FADS1*	rs174578	A/T	0.66	0.056(0.003)	1.86E-94
1-Arachidonoylglycerophosphoethanolamine*	Oily fish	Food	-0.097(0.015)	9.46E-11	*SLCO1B1*	rs4149056	T/C	0.84	-0.040(0.003)	3.66E-31
1-Eicosatrienoylglycerophosphocholine*	Oily fish	Food	-0.095(0.014)	1.12E-11	*FADS1*	rs968567	T/C	0.18	0.040(0.005)	2.84E-19
1-Eicosatrienoylglycerophosphocholine*	Other seafood	Food	-0.082(0.014)	8.44E-09	*FADS1*	rs968567	T/C	0.18	0.040(0.005)	2.84E-19
1-Linoleoylglycerophosphoethanolamine*	Oily fish	Food	-0.089(0.015)	1.60E-09	*FADS1*	rs174535	T/C	0.67	-0.044(0.004)	2.82E-36
10-Undecenoate (11:1n1)	Butter	Food	0.020(0.003)	2.35E-11	*CYP4A11*	rs9333029	A/G	0.87	-0.070(0.004)	1.52E-61

Table 10. SNP associations with diet-associated metabolites from previous genome-wide association study (Shin et al., Nat Genet. 2014 Jun;46(6):543-50)

			Diet variable-Metabolite (Meta-analysis result)						SNP-Metabolite (Shin et al., 2014)	
Metabolite	**Diet variable**	**Diet pattern or food**	**Beta(SE)**	***P***	**Gene symbol**	**SNP**	**EA/OA**	**EAF**	**Beta(SE)**	***P***
10-Undecenoate (11:1n1)	High Alcohol	Pattern	0.073(0.013)	8.17E-09	*CYP4A11*	rs9333029	A/G	0.87	-0.070(0.004)	1.52E-61
4-Androsten-3beta,17beta-diol disulfate 1*	Wine	Food	0.041(0.003)	1.50E-31	*SULT2A1*	rs296396	T/C	0.17	-0.175(0.009)	1.48E-92
4-Androsten-3beta,17beta-diol disulfate 1*	High Alcohol	Pattern	0.122(0.014)	3.27E-19	*SULT2A1*	rs296396	T/C	0.17	-0.175(0.009)	1.48E-92
4-Androsten-3beta,17beta-diol disulfate 1*	Spirits and liquors	Food	0.040(0.007)	4.30E-09	*SULT2A1*	rs296396	T/C	0.17	-0.175(0.009)	1.48E-92
5-Alpha-androstan-3beta,17beta-diol disulfate	Wine	Food	0.036(0.004)	2.71E-24	*CYP3A5*	rs10278040	A/G	0.04	-0.190(0.017)	1.17E-29
5-Alpha-androstan-3beta,17beta-diol disulfate	Wine	Food	0.036(0.004)	2.71E-24	*SULT2A1*	rs2547231	A/C	0.83	0.081(0.008)	3.35E-22
5-Alpha-androstan-3beta,17beta-diol disulfate	Wine	Food	0.036(0.004)	2.71E-24	*ZCWPW1*	rs13222543	T/C	0.02	-0.263(0.028)	1.18E-20
5-Alpha-androstan-3beta,17beta-diol disulfate	High Alcohol	Pattern	0.112(0.014)	1.73E-15	*CYP3A5*	rs10278040	A/G	0.04	-0.190(0.017)	1.17E-29
5-Alpha-androstan-3beta,17beta-diol disulfate	High Alcohol	Pattern	0.112(0.014)	1.73E-15	*SULT2A1*	rs2547231	A/C	0.83	0.081(0.008)	3.35E-22
5-Alpha-androstan-3beta,17beta-diol disulfate	High Alcohol	Pattern	0.112(0.014)	1.73E-15	*ZCWPW1*	rs13222543	T/C	0.02	-0.263(0.028)	1.18E-20
Docosapentaenoate (n3 DPA; 22:5n3)	Wine	Food	0.024(0.003)	5.28E-15	*FADS1*	rs174538	A/G	0.30	-0.026(0.004)	8.90E-14
Docosapentaenoate (n3 DPA; 22:5n3)	Other seafood	Food	0.089(0.013)	3.44E-11	*FADS1*	rs174538	A/G	0.30	-0.026(0.004)	8.90E-14
Docosapentaenoate (n3 DPA; 22:5n3)	Oily fish	Food	0.080(0.013)	2.22E-10	*FADS1*	rs174538	A/G	0.30	-0.026(0.004)	8.90E-14
Eicosapentaenoate (EPA; 20:5n3)	Oily fish	Food	0.169(0.014)	1.57E-33	*FADS1*	rs174556	T/C	0.30	-0.036(0.004)	1.97E-22
Eicosapentaenoate (EPA; 20:5n3)	Fruit & Vegetable	Pattern	0.095(0.009)	1.86E-25	*FADS1*	rs174556	T/C	0.30	-0.036(0.004)	1.97E-22
Eicosapentaenoate (EPA; 20:5n3)	Other seafood	Food	0.154(0.015)	2.66E-24	*FADS1*	rs174556	T/C	0.30	-0.036(0.004)	1.97E-22
Eicosapentaenoate (EPA; 20:5n3)	Wine	Food	0.027(0.003)	6.63E-22	*FADS1*	rs174556	T/C	0.30	-0.036(0.004)	1.97E-22

Table 10. SNP associations with diet-associated metabolites from previous genome-wide association study (Shin et al., Nat Genet. 2014 Jun;46(6):543-50)

			Diet variable-Metabolite (Meta-analysis result)						SNP-Metabolite (Shin et al., 2014)	
Metabolite	**Diet variable**	**Diet pattern or food**	**Beta(SE)**	***P***	**Gene symbol**	**SNP**	**EA/OA**	**EAF**	**Beta(SE)**	***P***
Eicosapentaenoate (EPA; 20:5n3)	Mediterranean Diet Score	Pattern	0.083(0.009)	9.54E-21	*FADS1*	rs174556	T/C	0.30	-0.036(0.004)	1.97E-22
Eicosapentaenoate (EPA; 20:5n3)	Low Meat	Pattern	-0.123(0.014)	8.44E-19	*FADS1*	rs174556	T/C	0.30	-0.036(0.004)	1.97E-22
Eicosapentaenoate (EPA; 20:5n3)	Baked goods	Food	-0.014(0.002)	1.08E-12	*FADS1*	rs174556	T/C	0.30	-0.036(0.004)	1.97E-22
Eicosapentaenoate (EPA; 20:5n3)	Avocado	Food	0.108(0.018)	2.01E-09	*FADS1*	rs174556	T/C	0.30	-0.036(0.004)	1.97E-22
Eicosapentaenoate (EPA; 20:5n3)	High fat salad dressings	Food	0.044(0.008)	5.48E-08	*FADS1*	rs174556	T/C	0.30	-0.036(0.004)	1.97E-22
Epiandrosterone sulfate	Wine	Food	0.018(0.003)	6.29E-09	*CYP3A5*	rs11974702	A/G	0.91	0.184(0.010)	2.80E-75
Epiandrosterone sulfate	Wine	Food	0.018(0.003)	6.29E-09	*ZCWPW1*	rs13222543	T/C	0.02	-0.347(0.024)	3.31E-47
Myo-inositol	Wine	Food	0.022(0.003)	1.36E-15	*SLC5A11*	rs4788439	T/C	0.08	-0.027(0.004)	6.59E-13
Myo-inositol	Fruit & Vegetable	Pattern	0.052(0.008)	4.43E-11	*SLC5A11*	rs4788439	T/C	0.08	-0.027(0.004)	6.59E-13
Nonanoylcarnitine*	Butter	Food	0.026(0.003)	6.48E-17	*ACADL*	rs3738934	T/C	0.62	-0.106(0.004)	1.21E-134
Nonanoylcarnitine*	Butter	Food	0.026(0.003)	6.48E-17	*THEM4*	rs12566232	A/C	0.71	-0.045(0.005)	5.69E-19
Scyllo-inositol	Wine	Food	0.052(0.003)	1.47E-49	*SLC5A11*	rs4787294	A/T	0.93	0.075(0.008)	9.64E-21
Scyllo-inositol	High Alcohol	Pattern	0.143(0.014)	4.20E-26	*SLC5A11*	rs4787294	A/T	0.93	0.075(0.008)	9.64E-21
Scyllo-inositol	Fruit & Vegetable	Pattern	0.088(0.009)	2.22E-23	*SLC5A11*	rs4787294	A/T	0.93	0.075(0.008)	9.64E-21
Scyllo-inositol	Baked goods	Food	-0.017(0.003)	1.44E-10	*SLC5A11*	rs4787294	A/T	0.93	0.075(0.008)	9.64E-21
Scyllo-inositol	Mediterranean Diet Score	Pattern	0.066(0.010)	2.25E-10	*SLC5A11*	rs4787294	A/T	0.93	0.075(0.008)	9.64E-21
Scyllo-inositol	Fried fish	Food	-0.154(0.026)	4.44E-09	*SLC5A11*	rs4787294	A/T	0.93	0.075(0.008)	9.64E-21
Stearidonate (18:4n3)	Wine	Food	0.019(0.003)	3.17E-10	*FADS1*	rs174601	T/C	0.34	-0.034(0.004)	7.93E-16
1-Methylxanthine	Coffee	Food	0.010(0.002)	3.31E-09	*NAT2*	rs4921914	T/C	0.78	0.088(0.005)	1.09E-60
X-01911	High Alcohol	Pattern	0.115(0.014)	3.77E-17	*COMT*	rs4680	A/G	0.51	-0.044(0.006)	1.31E-13
X-01911	Wine	Food	0.026(0.003)	2.55E-14	*COMT*	rs4680	A/G	0.51	-0.044(0.006)	1.31E-13
X-02249	Butter	Food	0.026(0.003)	2.31E-16	*CYP2C8*	rs1934955	A/G	0.71	-0.030(0.004)	3.33E-16
X-02249	Low Meat	Pattern	-0.110(0.015)	7.70E-13	*CYP2C8*	rs1934955	A/G	0.71	-0.030(0.004)	3.33E-16
X-02249	Dieting	Pattern	0.091(0.014)	4.01E-11	*CYP2C8*	rs1934955	A/G	0.71	-0.030(0.004)	3.33E-16
X-02269	Oily fish	Food	0.175(0.013)	1.44E-38	*CYP2C8*	rs2071426	T/C	0.71	0.055(0.007)	6.87E-14
X-02269	Fruit & Vegetable	Pattern	0.102(0.009)	7.46E-31	*CYP2C8*	rs2071426	T/C	0.71	0.055(0.007)	6.87E-14

Table 10. SNP associations with diet-associated metabolites from previous genome-wide association study (Shin et al., Nat Genet. 2014 Jun;46(6):543-50)

			Diet variable-Metabolite (Meta-analysis result)						SNP-Metabolite (Shin et al., 2014)	
Metabolite	**Diet variable**	**Diet pattern or food**	**Beta(SE)**	***P***	**Gene symbol**	**SNP**	**EA/OA**	**EAF**	**Beta(SE)**	***P***
X-02269	Other fish and seafood	Food	0.174(0.015)	9.25E-31	*CYP2C8*	rs2071426	T/C	0.71	0.055(0.007)	6.87E-14
X-02269	Mediterranean Diet Score	Pattern	0.099(0.009)	1.70E-29	*CYP2C8*	rs2071426	T/C	0.71	0.055(0.007)	6.87E-14
X-02269	Avocado	Food	0.147(0.019)	6.33E-15	*CYP2C8*	rs2071426	T/C	0.71	0.055(0.007)	6.87E-14
X-02269	Green leafy vegetables	Food	0.030(0.004)	2.39E-13	*CYP2C8*	rs2071426	T/C	0.71	0.055(0.007)	6.87E-14
X-02269	Low Meat	Pattern	-0.109(0.015)	2.61E-13	*CYP2C8*	rs2071426	T/C	0.71	0.055(0.007)	6.87E-14
X-02269	High fibre breakfast cereals	Food	0.030(0.004)	3.63E-12	*CYP2C8*	rs2071426	T/C	0.71	0.055(0.007)	6.87E-14
X-02269	Allium vegetables	Food	0.023(0.004)	3.55E-08	*CYP2C8*	rs2071426	T/C	0.71	0.055(0.007)	6.87E-14
X-08402	Butter	Food	0.017(0.003)	7.17E-10	*SGPP1*	rs7157785	T/G	0.16	0.068(0.003)	7.23E-87
X-08402	Butter	Food	0.017(0.003)	7.17E-10	*SPTLC3*	rs4814176	T/C	0.39	0.022(0.003)	2.97E-17
X-09789	Porridge	Food	0.094(0.008)	4.96E-33	*SLC51A*	rs7642243	C/G	0.39	0.048(0.006)	7.5E-16
X-09789	High Alcohol	Pattern	-0.116(0.013)	4.99E-19	*SLC51A*	rs7642243	C/G	0.39	0.048(0.006)	7.50E-16
X-09789	High fibre breakfast cereals	Food	0.038(0.005)	7.28E-14	*SLC51A*	rs7642243	C/G	0.39	0.048(0.006)	7.5E-16
X-09789	Fruit & Vegetable	Pattern	0.061(0.009)	3.50E-11	*SLC51A*	rs7642243	C/G	0.39	0.048(0.006)	7.50E-16
X-09789	Apples & pears	Food	0.020(0.004)	2.22E-08	*SLC51A*	rs7642243	C/G	0.39	0.048(0.006)	7.5E-16
X-09789	Wholemeal bread, grains	Food	0.015(0.003)	6.03E-08	*SLC51A*	rs7642243	C/G	0.39	0.048(0.006)	7.5E-16
X-10395	Wine	Food	0.031(0.003)	3.45E-26	*NR1I3*	rs4073054	A/C	0.63	0.016(0.002)	4.42E-19
X-10395	High Alcohol	Pattern	0.091(0.013)	3.80E-13	*NR1I3*	rs4073054	A/C	0.63	0.016(0.002)	4.42E-19
X-10510	Butter	Food	0.022(0.003)	8.28E-15	*SGPP1*	rs7157785	T/G	0.86	-0.044(0.004)	3.41E-37
X-11261	Low Meat	Pattern	0.105(0.014)	4.20E-14	*SLC22A1*	rs662138	C/G	0.83	0.063(0.006)	9.38E-25
X-11261	Low Meat	Pattern	0.105(0.014)	4.20E-14	*SLC16A9*	rs1171614	T/C	0.22	-0.042(0.006)	6.68E-14
X-11261	Fruit & Vegetable	Pattern	-0.061(0.010)	2.08E-10	*SLC22A1*	rs662138	C/G	0.83	0.063(0.006)	9.38E-25
X-11261	Fruit & Vegetable	Pattern	-0.061(0.010)	2.08E-10	*SLC16A9*	rs1171614	T/C	0.22	-0.042(0.006)	6.68E-14
X-11315	Fruit & Vegetable	Pattern	0.142(0.008)	3.87E-63	*SLC6A20*	rs4327428	A/C	0.11	-0.031(0.004)	2.81E-12
X-11315	Mediterranean Diet Score	Pattern	0.109(0.009)	7.94E-34	*SLC6A20*	rs4327428	A/C	0.11	-0.031(0.004)	2.81E-12
X-11315	Nuts	Food	0.054(0.005)	3.75E-25	*SLC6A20*	rs4327428	A/C	0.11	-0.031(0.004)	2.81E-12
X-11315	Apples & pears	Food	0.035(0.004)	9.63E-20	*SLC6A20*	rs4327428	A/C	0.11	-0.031(0.004)	2.81E-12

Table 10. SNP associations with diet-associated metabolites from previous genome-wide association study (Shin et al., Nat Genet. 2014 Jun;46(6):543-50)

			Diet variable-Metabolite (Meta-analysis result)						SNP-Metabolite (Shin et al., 2014)	
Metabolite	**Diet variable**	**Diet pattern or food**	**Beta(SE)**	***P***	**Gene symbol**	**SNP**	**EA/OA**	**EAF**	**Beta(SE)**	***P***
X-11315	Oily fish	Food	0.106(0.014)	1.62E-14	*SLC6A20*	rs4327428	A/C	0.11	-0.031(0.004)	2.81E-12
X-11315	Berries	Food	0.103(0.013)	2.70E-14	*SLC6A20*	rs4327428	A/C	0.11	-0.031(0.004)	2.81E-12
X-11315	Peaches	Food	0.127(0.017)	5.92E-14	*SLC6A20*	rs4327428	A/C	0.11	-0.031(0.004)	2.81E-12
X-11315	Green leafy vegetables	Food	0.033(0.005)	1.71E-12	*SLC6A20*	rs4327428	A/C	0.11	-0.031(0.004)	2.81E-12
X-11315	White and brown bread, refined grains	Food	-0.017(0.002)	1.84E-12	*SLC6A20*	rs4327428	A/C	0.11	-0.031(0.004)	2.81E-12
X-11315	Confectionary & jams	Food	-0.010(0.001)	1.53E-11	*SLC6A20*	rs4327428	A/C	0.11	-0.031(0.004)	2.81E-12
X-11315	Traditional English	Pattern	-0.093(0.014)	1.68E-11	*SLC6A20*	rs4327428	A/C	0.11	-0.031(0.004)	2.81E-12
X-11315	Savoury pies	Food	-0.221(0.034)	9.08E-11	*SLC6A20*	rs4327428	A/C	0.11	-0.031(0.004)	2.81E-12
X-11315	Citrus fruit	Food	0.026(0.004)	6.50E-10	*SLC6A20*	rs4327428	A/C	0.11	-0.031(0.004)	2.81E-12
X-11315	Fried fish	Food	-0.164(0.027)	1.54E-09	*SLC6A20*	rs4327428	A/C	0.11	-0.031(0.004)	2.81E-12
X-11315	High fibre breakfast cereals	Food	0.030(0.005)	2.75E-09	*SLC6A20*	rs4327428	A/C	0.11	-0.031(0.004)	2.81E-12
X-11315	Baked goods	Food	-0.012(0.002)	2.90E-08	*SLC6A20*	rs4327428	A/C	0.11	-0.031(0.004)	2.81E-12
X-11381	Mediterranean Diet Score	Pattern	-0.069(0.009)	8.27E-14	*SLC16A9*	rs12356193	A/G	0.84	0.024(0.003)	8.36E-20
X-11381	High Alcohol	Pattern	-0.077(0.012)	4.54E-11	*SLC16A9*	rs12356193	A/G	0.84	0.024(0.003)	8.36E-20
X-11381	Processed meats	Food	0.049(0.008)	4.74E-10	*SLC16A9*	rs12356193	A/G	0.84	0.024(0.003)	8.36E-20
X-11381	Traditional English	Pattern	0.072(0.012)	3.71E-09	*SLC16A9*	rs12356193	A/G	0.84	0.024(0.003)	8.36E-20
X-11381	Soy foods	Food	-0.108(0.020)	5.80E-08	*SLC16A9*	rs12356193	A/G	0.84	0.024(0.003)	8.36E-20
X-11469	Oily fish	Food	0.176(0.013)	5.87E-39	*CYP2C8*	rs2071426	T/C	0.71	0.052(0.007)	3.97E-14
X-11469	Oily fish	Food	0.176(0.013)	5.87E-39	*SLC17A3*	rs11754288	A/G	0.44	0.043(0.006)	1.88E-12
X-11469	Fruit & Vegetable	Pattern	0.106(0.009)	3.39E-34	*CYP2C8*	rs2071426	T/C	0.71	0.052(0.007)	3.97E-14
X-11469	Other fish and seafood	Food	0.175(0.015)	2.99E-31	*CYP2C8*	rs2071426	T/C	0.71	0.052(0.007)	3.97E-14
X-11469	Other fish and seafood	Food	0.175(0.015)	2.99E-31	*SLC17A3*	rs11754288	A/G	0.44	0.043(0.006)	1.88E-12
X-11469	Mediterranean Diet Score	Pattern	0.100(0.009)	9.16E-31	*CYP2C8*	rs2071426	T/C	0.71	0.052(0.007)	3.97E-14
X-11469	Avocado	Food	0.153(0.019)	9.36E-16	*CYP2C8*	rs2071426	T/C	0.71	0.052(0.007)	3.97E-14
X-11469	Avocado	Food	0.153(0.019)	9.36E-16	*SLC17A3*	rs11754288	A/G	0.44	0.043(0.006)	1.88E-12

Table 10. SNP associations with diet-associated metabolites from previous genome-wide association study (Shin et al., Nat Genet. 2014 Jun;46(6):543-50)

			Diet variable-Metabolite (Meta-analysis result)						SNP-Metabolite (Shin et al., 2014)	
Metabolite	**Diet variable**	**Diet pattern or food**	**Beta(SE)**	***P***	**Gene symbol**	**SNP**	**EA/OA**	**EAF**	**Beta(SE)**	***P***
X-11469	Green leafy vegetables	Food	0.030(0.004)	5.66E-14	*CYP2C8*	rs2071426	T/C	0.71	0.052(0.007)	3.97E-14
X-11469	Green leafy vegetables	Food	0.030(0.004)	5.66E-14	*SLC17A3*	rs11754288	A/G	0.44	0.043(0.006)	1.88E-12
X-11469	Low Meat	Pattern	-0.110(0.015)	1.88E-13	*SLC17A3*	rs11754288	A/G	0.44	0.043(0.006)	1.88E-12
X-11469	High fibre breakfast cereals	Food	0.031(0.004)	5.40E-13	*CYP2C8*	rs2071426	T/C	0.71	0.052(0.007)	3.97E-14
X-11469	High fibre breakfast cereals	Food	0.031(0.004)	5.40E-13	*SLC17A3*	rs11754288	A/G	0.44	0.043(0.006)	1.88E-12
X-11478	Low Meat	Pattern	0.094(0.015)	9.86E-11	*ACSM2A*	rs6497490	T/G	0.88	-0.082(0.008)	4.95E-27
X-11550	Wine	Food	0.018(0.003)	3.94E-11	*CETP*	rs247616	T/C	0.32	0.010(0.001)	1.11E-11
X-11799	Wine	Food	0.033(0.003)	1.66E-27	*GBA3*	rs3099557	A/G	0.85	-0.148(0.011)	1.83E-39
X-11799	High Alcohol	Pattern	0.120(0.014)	6.07E-18	*GBA3*	rs3099557	A/G	0.85	-0.148(0.011)	1.83E-39
X-11799	Baked goods	Food	-0.017(0.003)	4.88E-09	*GBA3*	rs3099557	A/G	0.85	-0.148(0.011)	1.83E-39
X-12038	Wine	Food	0.022(0.003)	9.20E-13	*CETP*	rs1800775	A/C	0.50	0.013(0.002)	1.02E-11
X-12038	High Alcohol	Pattern	0.064(0.012)	1.02E-07	*CETP*	rs1800775	A/C	0.50	0.013(0.002)	1.02E-11
X-12063	Fruit & Vegetable	Pattern	-0.048(0.008)	2.31E-09	*CYP3A5*	rs10242455	A/G	0.93	0.221(0.010)	1.67E-109
X-12063	Fruit & Vegetable	Pattern	-0.048(0.008)	2.31E-09	*SLCO1B1*	rs4149056	T/C	0.84	-0.118(0.007)	1.26E-73
X-12627	Wine	Food	0.022(0.003)	5.88E-15	*ELOVL2*	rs4713169	C/G	0.42	-0.032(0.004)	2.87E-14
X-12627	Oily fish	Food	0.073(0.012)	5.74E-09	*ELOVL2*	rs4713169	C/G	0.42	-0.032(0.004)	2.87E-14
X-12798	Low fat milk	Food	0.062(0.008)	1.24E-15	*SLC22A1*	rs316019	A/C	0.09	-0.181(0.005)	1.56E-259
X-12798	Low fat milk	Food	0.062(0.008)	1.24E-15	*SLC16A9*	rs1171615	T/C	0.78	-0.048(0.004)	3.19E-34
X-12798	High Alcohol	Pattern	-0.099(0.012)	1.58E-15	*SLC22A1*	rs316019	A/C	0.09	-0.181(0.005)	1.56E-259
X-12798	Mediterranean Diet Score	Pattern	-0.069(0.010)	6.80E-13	*SLC16A9*	rs1171615	T/C	0.78	-0.048(0.004)	3.19E-34
X-13477	Mediterranean Diet Score	Pattern	0.067(0.010)	1.56E-10	*NAT8*	rs10206899	T/C	0.78	0.025(0.003)	1.73E-14
X-14473	Coffee	Food	0.038(0.001)	6.12E-187	*CYP2C9*	rs4986894	T/C	0.84	-0.042(0.006)	5.55E-12
X-14473	Black tea	Food	-0.024(0.001)	1.36E-72	*CYP2C9*	rs4986894	T/C	0.84	-0.042(0.006)	5.55E-12
X-14473	Dieting	Pattern	-0.148(0.014)	8.46E-25	*CYP2C9*	rs4986894	T/C	0.84	-0.042(0.006)	5.55E-12
X-21365 [trimethyl-N-aminovalerate]	Low fat milk	Food	0.076(0.007)	9.36E-27	*SLC22A4*	rs273913	T/C	0.38	0.026(0.003)	1.08E-25

Table 10. SNP associations with diet-associated metabolites from previous genome-wide association study (Shin et al., Nat Genet. 2014 Jun;46(6):543-50)

			Diet variable-Metabolite (Meta-analysis result)						SNP-Metabolite (Shin et al., 2014)	
Metabolite	**Diet variable**	**Diet pattern or food**	**Beta(SE)**	***P***	**Gene symbol**	**SNP**	**EA/OA**	**EAF**	**Beta(SE)**	***P***
X-21365 [trimethyl-N-aminovalerate]	Low fat milk	Food	0.076(0.007)	9.36E-27	*MARCH8*	rs2291429	A/C	0.76	0.017(0.003)	8.69E-11
X-21365 [trimethyl-N-aminovalerate]	Mediterranean Diet Score	Pattern	-0.073(0.009)	1.48E-15	*SLC22A4*	rs273913	T/C	0.38	0.026(0.003)	1.08E-25
X-21365 [trimethyl-N-aminovalerate]	Mediterranean Diet Score	Pattern	-0.073(0.009)	1.48E-15	*MARCH8*	rs2291429	A/C	0.76	0.017(0.003)	8.69E-11
X-21365 [trimethyl-N-aminovalerate]	High Alcohol	Pattern	-0.091(0.012)	3.91E-15	*SLC22A4*	rs273913	T/C	0.38	0.026(0.003)	1.08E-25
X-21365 [trimethyl-N-aminovalerate]	High Alcohol	Pattern	-0.091(0.012)	3.91E-15	*MARCH8*	rs2291429	A/C	0.76	0.017(0.003)	8.69E-11

Notes: Statistics are reported for the most associated metabolite at each single nucleotide polymorphism (SNP) for those metabolites associated with food intakes and diet patterns. The genome-wide association study was previously performed on the TwinsUK and KORA blood metabolite datasets (Shin et al., Nat Genet. 2014 Jun;46(6):543-50). EA/OA = effect/other allele; EAF = effect allele frequency.

Table 11. Diet-SNP associations with SNPs associated with metabolites from a previous genome-wide association study (Shin et al., Nat Genet. 2014 Jun;46(6):543-50)

			Food Group-Metabolite (Meta-analysis result)			SNP-Metabolite (Shin et al., 2014)		Food variable-SNP		
Metabolite	**Diet variable**	**Diet pattern or food**	**Beta(SE)**	***P***	**SNP**	**Beta(SE)**	***P***	**Beta(SE)**	***P***	**Gene symbol**
3-Phenylpropionate (hydrocinnamate)	Fruit & Vegetable	Pattern	0.080(0.010)	1.73E-16	rs11647589	0.030(0.005)	2.85E-11	0.140(0.059)	1.80E-02	*ACSM5*
3-Phenylpropionate (hydrocinnamate)	Savoury pies	Food	-0.231(0.037)	3.72E-10	rs11647589	0.030(0.005)	2.85E-11	-0.023(0.012)	4.76E-02	*ACSM5*
Indolepropionate	Fruid & Vegetable	Pattern	0.082(0.009)	8.42E-21	rs1394678	-0.035(0.004)	1.70E-20	0.163(0.065)	1.20E-02	*ACSM2A*
Indolepropionate	Mediterranean Diet Score	Pattern	0.078(0.009)	1.60E-17	rs1394678	-0.035(0.004)	1.70E-20	0.125(0.053)	1.80E-02	*ACSM2A*
Pyroglutamine*	Meat	Food	-0.062(0.010)	2.10E-10	rs17279437	0.059(0.006)	1.25E-20	-0.142(0.069)	3.95E-02	*SLC6A20*
1,5-Anhydroglucitol (1,5-AG)	Other seafood	Food	-0.081(0.014)	1.80E-08	rs3800993	0.029(0.004)	1.53E-15	-0.090(0.034)	8.51E-03	*MGAM*
5-Alpha-androstan-3beta,17beta-diol disulfate	High Alcohol	Pattern	0.112(0.014)	1.73E-15	rs10278040	-0.190(0.017)	1.17E-29	0.198(0.077)	1.10E-02	*CYP3A5*
5-Alpha-androstan-3beta,17beta-diol disulfate	High Alcohol	Pattern	0.112(0.014)	1.73E-15	rs13222543	-0.263(0.028)	1.18E-20	0.301(0.129)	2.00E-02	*ZCWPW1*
Myo-inositol	Wine	Food	0.022(0.003)	1.36E-15	rs4788439	-0.027(0.004)	6.59E-13	-0.654(0.236)	5.59E-03	*SLC5A11*
Scyllo-inositol	Fried fish	Food	-0.154(0.026)	4.44E-09	rs4787294	0.075(0.008)	9.64E-21	0.088(0.035)	1.25E-02	*SLC5A11*
Scyllo-inositol	Wine	Food	0.052(0.003)	1.47E-49	rs4787294	0.075(0.008)	9.64E-21	-0.586(0.243)	1.60E-02	*SLC5A11*
X-02269	Green leafy vegetables	Food	0.030(0.004)	2.39E-13	rs2071426	0.055(0.007)	6.87E-14	0.228(0.110)	3.76E-02	*CYP2C8*
X-08402	Butter	Food	0.017(0.003)	7.17E-10	rs7157785	0.068(0.003)	7.23E-87	0.004(0.001)	2.95E-03	*SGPP1*
X-10395	High Alcohol	Pattern	0.091(0.013)	3.80E-13	rs4073054	0.016(0.002)	4.42E-19	-0.071(0.034)	4.10E-02	*NR1I3*
X-10510	Butter	Food	0.022(0.003)	8.28E-15	rs7157785	-0.044(0.004)	3.41E-37	0.004(0.001)	2.95E-03	*SGPP1*
X-11315	Citrus fruit	Food	0.026(0.004)	6.50E-10	rs4327428	-0.031(0.004)	2.81E-12	0.473(0.209)	2.37E-02	*SLC6A20*
X-11315	Traditional English	Pattern	-0.093(0.014)	1.68E-11	rs4327428	-0.031(0.004)	2.81E-12	-0.114(0.057)	4.80E-02	*SLC6A20*
X-11469	Green leafy vegetables	Food	0.030(0.004)	5.66E-14	rs2071426	0.052(0.007)	3.97E-14	0.228(0.110)	3.76E-02	*CYP2C8*
X-11469	Avocado	Food	0.153(0.019)	9.36E-16	rs11754288	0.043(0.006)	1.88E-12	0.043(0.022)	4.75E-02	*SLC17A3*

Notes: Statistics are reported for associations between food intake at metabolite-SNPs which meet the 5% level of significance. A linear regression was performed using the applicable metabolite-associated SNP as a predictor of reported food group intake or dietary pattern score adjusted for age, energy intake and family relatedness; an additive genetic model was performed for each. No associations met statistical significance of 4.76×10^{-4} (0.05/105 tests). The genome-wide association study was previously performed on the TwinsUK and KORA blood metabolite datasets (Shin et al., Nat Genet. 2014 Jun;46(6):543-50).

Appendix D. Chapter 5 Appendices

Document 1: Food Preference Questionnaire

Your food and lifestyle preferences

Welcome to our latest questionnaire.

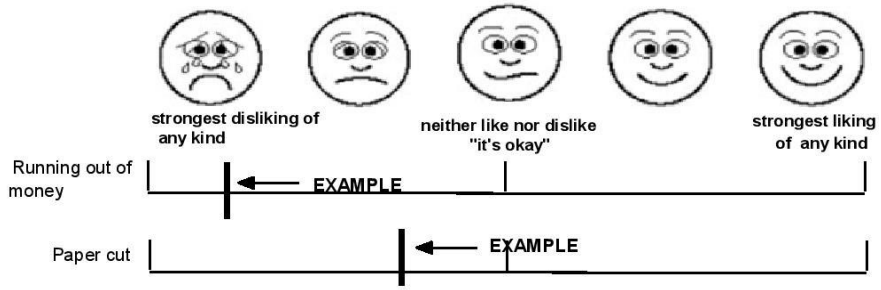
Throughout this section of the questionnaire you will use the scales shown below to report your likes and dislikes for foods, physical activities, and experiences.

The term **'Dislike'** ranges from weak disliking (close to neutral) to strongest disliking of any kind (left-hand side of the scale).
The term **'Like'** is right of neutral, ranging from weaker liking (close to neutral) to strongest liking of any kind (right-hand side of the scale).
The term **'Neutral'** is in the middle (neither like nor dislike; "it's okay").

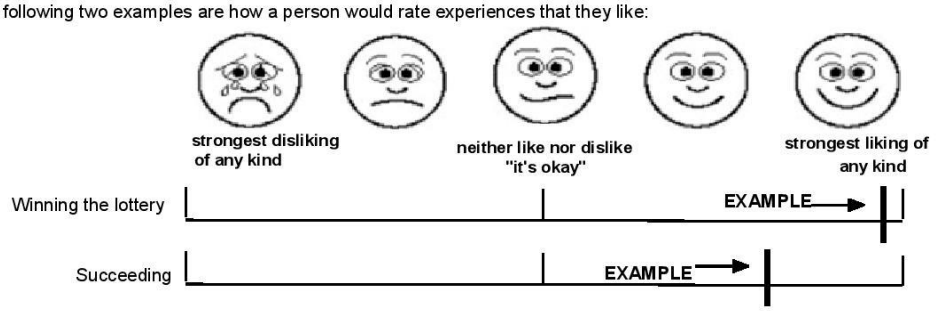
Questions marked with (*) are mandatory.

Below are some examples to show you how to use the scales.

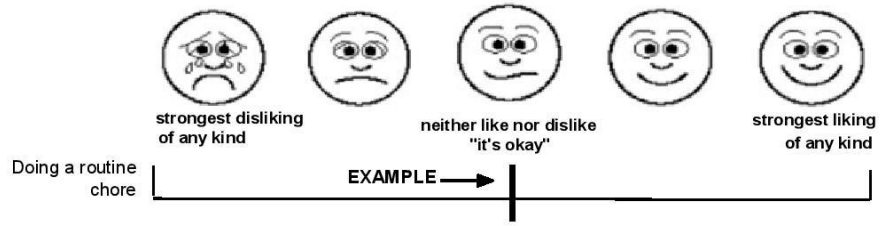
The following two examples are how a person would rate experiences that they dislike:



The following two examples are how a person would rate experiences that they like:



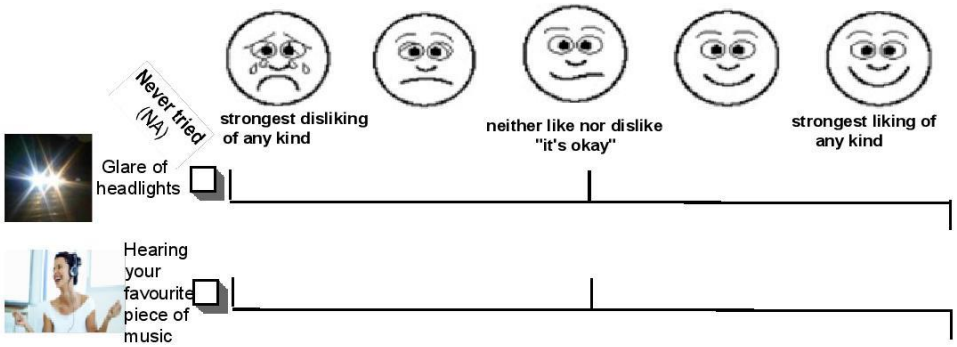
The following example is how a person would rate an experience that they neither like nor dislike (a neutral experience):



* Please rate your liking or disliking of the following.

The ends of the scale are the strongest disliking or liking of any kind. You can place your rating anywhere between the labels.

To place your rating on the scale, either click the position on the scale that you would like to choose, or drag the slider to that place. If your rating is neutral, simply click on the slider in its starting position at the center of the scale. If you have never tried a food, physical activity or experience, then select the "never tried" option.



Use the scales below to report your likes and dislikes for foods, physical activities, and experiences.

To place your rating on the scale, either click the position on the scale that you would like to choose, or drag the slider to that place. If your rating is neutral, simply click on the slider in its starting position at the center of the scale. If you have never tried a food, physical activity or experience, then select the "never tried" option.

	Never tried (NA)	strongest disliking of any kind		neither like nor dislike		strongest liking of any kind
 broccoli	☐					
 tortilla chips or crisps	☐					
 crispy bacon	☐					
 beer	☐					
 prawns & shellfish	☐					
 butter/ margarine	☐					
 chips	☐					
 wholemeal bread	☐					
 blue cheese	☐					
 skimmed milk	☐					

Page 3

Use the scales below to report your likes and dislikes for foods, physical activities, and experiences.

To place your rating on the scale, either click the position on the scale that you would like to choose, or drag the slider to that place. If your rating is neutral, simply click on the slider in its starting position at the center of the scale. If you have never tried a food, physical activity or experience, then select the "never tried" option.


fresh tomatoes


black coffee


cooling off on a hot day


pizza


plain yoghurt


fried chicken


burn of spicy foods


bagel/ rolls


sausage


soy sauce

Page 4

Use the scales below to report your likes and dislikes for foods, physical activities, and experiences.

To place your rating on the scale, either click the position on the scale that you would like to choose, or drag the slider to that place. If your rating is neutral, simply click on the slider in its starting position at the center of the scale. If you have never tried a food, physical activity or experience, then select the "never tried" option.


taking the stairs


fizzy soft drinks & sweet drinks (e.g. lemonade, squashes)


salty pretzels


fresh coriander


pasta/ noodles


going to the gym


mayonnaise


chili pepper


whole milk


baked chicken

Use the scales below to report your likes and dislikes for foods, physical activities, and experiences.

To place your rating on the scale, either click the position on the scale that you would like to choose, or drag the slider to that place. If your rating is neutral, simply click on the slider in its starting position at the center of the scale. If you have never tried a food, physical activity or experience, then select the "never tried" option.


olives ☐


dark chocolate ☐


garlic ☐


ketchup ☐


ice cream ☐


grapefruit ☐


corn flakes ☐

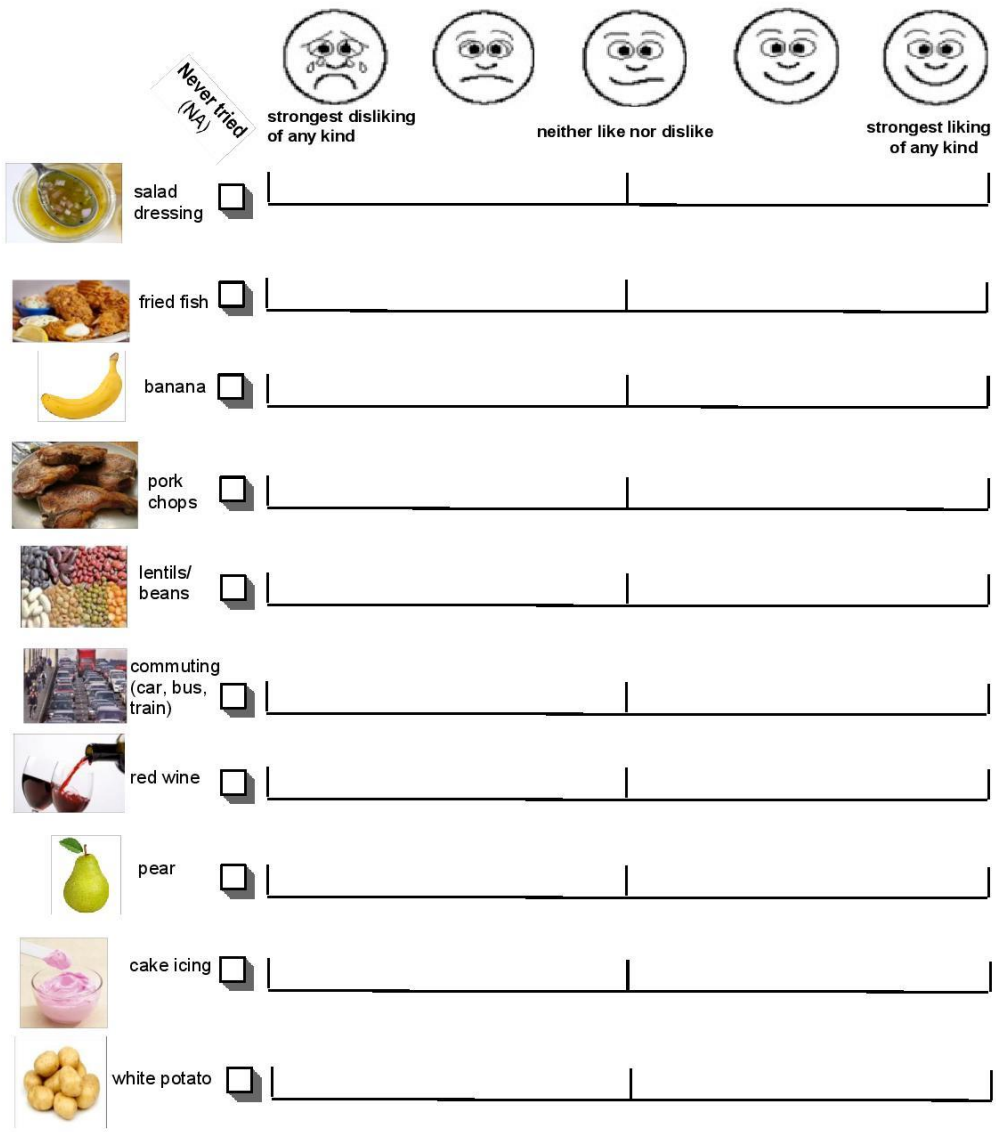

exercising alone ☐


white rice ☐


Smell of freshly cut grass ☐

Use the scales below to report your likes and dislikes for foods, physical activities, and experiences.

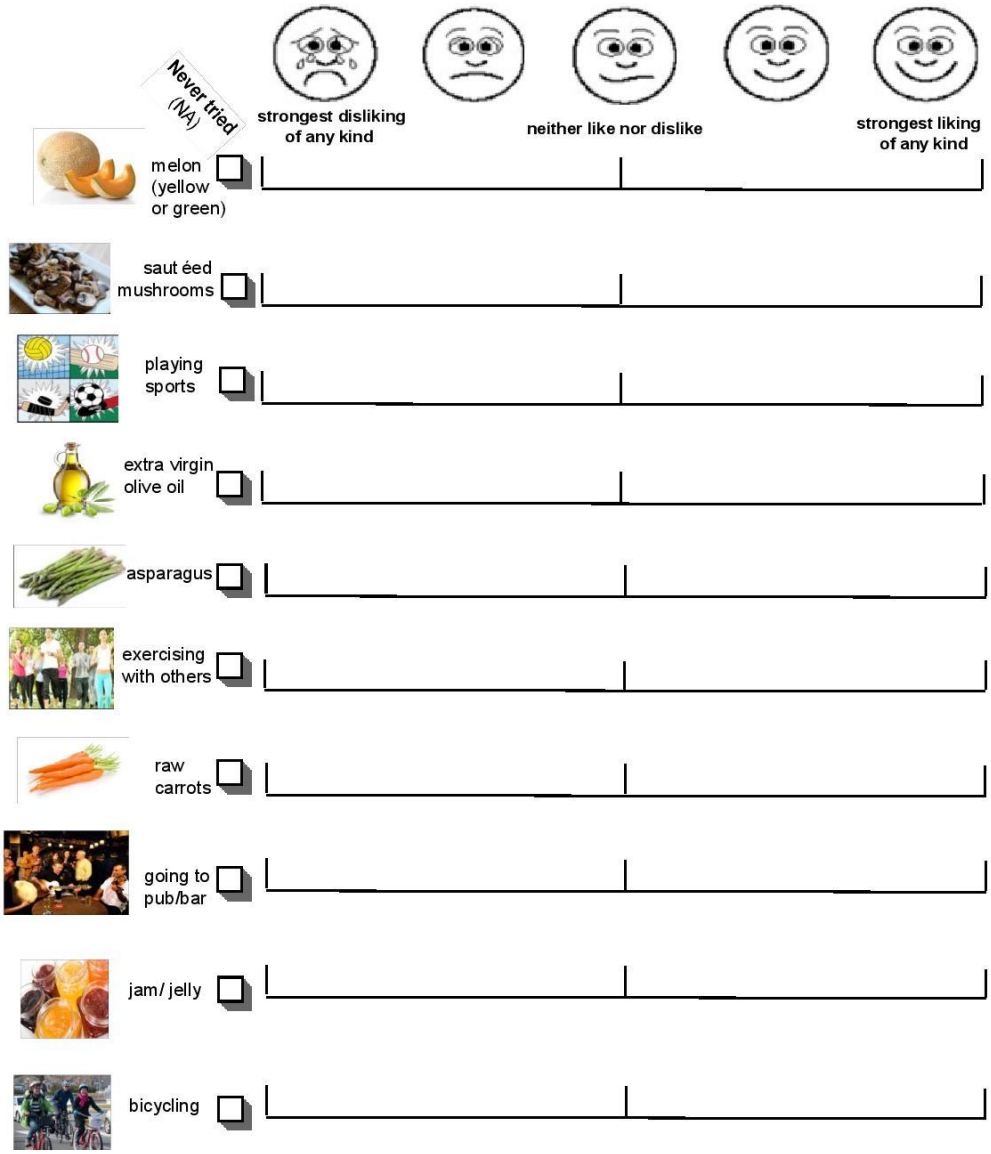
To place your rating on the scale, either click the position on the scale that you would like to choose, or drag the slider to that place. If your rating is neutral, simply click on the slider in its starting position at the center of the scale. If you have never tried a food, physical activity or experience, then select the "never tried" option.



Page 7

Use the scales below to report your likes and dislikes for foods, physical activities, and experiences.

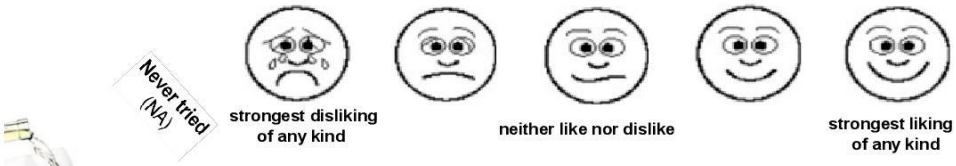
To place your rating on the scale, either click the position on the scale that you would like to choose, or drag the slider to that place. If your rating is neutral, simply click on the slider in its starting position at the center of the scale. If you have never tried a food, physical activity or experience, then select the "never tried" option.



Page 8

Use the scales below to report your likes and dislikes for foods, physical activities, and experiences.

To place your rating on the scale, either click the position on the scale that you would like to choose, or drag the slider to that place. If your rating is neutral, simply click on the slider in its starting position at the center of the scale. If you have never tried a food, physical activity or experience, then select the "never tried" option.




white wine ☐


watching television ☐


lemon ☐


porridge ☐


cherries ☐


vinegar ☐


coffee or tea with sugar ☐


beetroot ☐


horseradish/ wasabi ☐


working up a sweat ☐

Page 9

Use the scales below to report your likes and dislikes for foods, physical activities, and experiences.

To place your rating on the scale, either click the position on the scale that you would like to choose, or drag the slider to that place. If your rating is neutral, simply click on the slider in its starting position at the center of the scale. If you have never tried a food, physical activity or experience, then select the "never tried" option.


vodka, gin, scotch ☐


tuna or salmon ☐


getting caught in a lie ☐


biscuits, cakes or pastries ☐


ham ☐


high-fibre bar ☐


strawberries ☐


savoury biscuits ☐


gherkins ☐


seeing a mouse at home ☐

Use the scales below to report your likes and dislikes for foods, physical activities, and experiences.

To place your rating on the scale, either click the position on the scale that you would like to choose, or drag the slider to that place. If your rating is neutral, simply click on the slider in its starting position at the center of the scale. If you have never tried a food, physical activity or experience, then select the "never tried" option.

Never tried (NA)	strongest disliking of any kind		neither like nor dislike		strongest liking of any kind
 spinach/ greens	☐				
 pineapple	☐				
 hot tea	☐				
 cheese cake	☐				
 salting food	☐				
 char-grilled meats	☐				
 aubergine	☐				
 cheddar cheese	☐				
 Tabasco sauce	☐				
 raw onion	☐				

Page 11

Use the scales below to report your likes and dislikes for foods, physical activities, and experiences.

To place your rating on the scale, either click the position on the scale that you would like to choose, or drag the slider to that place. If your rating is neutral, simply click on the slider in its starting position at the center of the scale. If you have never tried a food, physical activity or experience, then select the "never tried" option.

Never tried (NA)

strongest disliking of any kind

neither like nor dislike

strongest liking of any kind

going to a café

orange juice

sweet coffee drinks & whipped cream

black pepper

cigarette smoking

faux meat products (e.g. Quorn)

curries

diet fizzy soft drinks (e.g. Coke Zero, Diet Coke)

unsalted nuts

beef steak

Page 12

Use the scales below to report your likes and dislikes for foods, physical activities, and experiences.

To place your rating on the scale, either click the position on the scale that you would like to choose, or drag the slider to that place. If your rating is neutral, simply click on the slider in its starting position at the center of the scale. If you have never tried a food, physical activity or experience, then select the "never tried" option.

eggs ☐

soy milk ☐

low calorie sweetener in coffee, teas or other beverages (e.g. Canderel) ☐

Food avoidance questions:

A			B								
Do you avoid any of the items listed below?			**If yes, what is/are the reason(s) why you avoid this food? (select all that apply)**								
			(1) I am allergic	(2) I am intolerant (e.g. Impaired digestion of the food)	(3) I am sensitive (e.g. I feel discomfort following eating)	(4) I have an illness/ disease (e.g. Inflammatory bowel disease)	(5) I believe it is better for long-term health (e.g. reduced risk of chronic disease)	(6) I believe it is better for weight loss/ maintenance	(7) Because of my beliefs (e.g. religion, veganism)	(8) I dislike this food	(9) Other
	(0) No	(1) Yes									
1. Fish											
2. Shellfish											
3. Red meat											
4. Poultry											
5. Gluten											
6. Dairy											
7. Eggs											
8. Sweets											
9. Soy											
10. Nuts											
11. Alcohol											

2. Do you avoid any other foods?

(0) No (skip to question 3)	(1) Yes, please complete the table below..

A		B								
What other foods do you avoid?		**If yes, what is/are the reason(s) why you avoid this food? (select all that apply)**								
Please click on the box and type only one food type per box.		(1) I am allergic	(2) I am intolerant (e.g. Impaired digestion of the food)	(3) I am sensitive (e.g. I feel discomfort following eating)	(4) I have an illness/ disease (e.g. Inflammatory bowel disease)	(5) I believe it is better for long-term health (e.g. reduced risk of chronic disease)	(6) I believe it is better for weight loss/ maintenance	(7) Because of my beliefs (e.g. religion, veganism)	(8) I dislike this food	(9) Other
12. Other:										
13. Other:										
14. Other:										
15. Other:										
16. Other:										

Eating rate questions:

3. How would you compare your eating rate to others?

(1) ☐ very slow	(2) ☐ relatively slow	(3) ☐ medium	(4) ☐ relatively fast	(5) ☐ very fast

4. Do you usually eat until you feel full?

(1) ☐ I quit eating before I feel full	(2) ☐ I quit eating when I feel full	(3) ☐ I continue eating after I feel full

5. Between you and your twin: Who usually eats the most?

(1) ☐ Me	(2) ☐ We eat about the same amount	(3) ☐ My twin	(9999902) ☐ Don't know

Press SUBMIT to complete your questionnaire, otherwise, press BACK button to modify your answers.

THANK YOU VERY MUCH FOR TAKING THE TIME TO COMPLETE THIS QUESTIONNAIRE. WE APPRECIATE ALL YOUR HELP

Table 1. Training group food group metabolite associations

Food group	Metabolite	Super-pathway	Sub-pathway	Discovery		Discordant		Meta-analysis	
				Beta (SE)	*P*	*Beta (SE)*	*P*	*Beta (SE)*	*P*
Vegetables	X-11315			0.009 (0.002)	1.06E-08	0.007 (0.005)	1.51E-01	0.009 (0.002)	2.94E-09
	X-11372			-0.007 (0.001)	1.09E-06	-0.004 (0.004)	3.06E-01	-0.007 (0.001)	6.20E-07
	X-12063			-0.006 (0.001)	2.36E-06	-0.002 (0.003)	4.28E-01	-0.005 (0.001)	2.33E-06
Fruit	**stachydrine**	Xenobiotics	Food component, Plant	0.021 (0.002)	3.65E-17	0.012 (0.004)	6.86E-03	0.019 (0.002)	8.71E-19
	glycerate	Carbohydrate	Glycolysis, gluconeogenesis, pyruvate metabolism	0.017 (0.002)	7.92E-13	0.017 (0.004)	2.33E-04	0.017 (0.002)	2.08E-16
	threonate	Cofactors and vitamins	Ascorbate and aldarate metabolism	0.012 (0.002)	2.47E-08	0.015 (0.004)	3.38E-04	0.013 (0.002)	1.89E-11
	threitol	Carbohydrate	Nucleotide sugars, pentose metabolism	0.012 (0.002)	5.18E-10	0.005 (0.003)	1.19E-01	0.010 (0.002)	7.27E-10
	docosahexaenoate (DHA; 22:6n3)	Lipid	Essential fatty acid	0.011 (0.002)	6.54E-06	0.010 (0.004)	2.13E-02	0.011 (0.002)	3.42E-07
	scyllo-inositol	Lipid	Inositol metabolism	0.012 (0.002)	1.03E-06	0.007 (0.004)	1.11E-01	0.011 (0.002)	3.68E-07
	proline	Amino acid	Urea cycle; arginine-, proline-, metabolism	-0.010 (0.002)	1.70E-06	-0.005 (0.005)	3.33E-01	-0.009 (0.002)	1.30E-06
	hippurate	Xenobiotics	Benzoate metabolism	0.011 (0.002)	7.86E-07	0.003 (0.004)	4.69E-01	0.009 (0.002)	2.12E-06
	X-11315			0.021 (0.002)	1.08E-16	0.014 (0.004)	1.43E-03	0.019 (0.002)	2.52E-19
	X-11372			-0.012 (0.002)	5.68E-08	-0.007 (0.004)	8.31E-02	-0.011 (0.002)	1.66E-08
Whole grains	3-phenylpropionate (hydrocinnamate)	Amino acid	Phenylalanine & tyrosine metabolism	0.017 (0.003)	1.90E-07	0.003 (0.007)	6.79E-01	0.015 (0.003)	7.90E-07
	eicosapentaenoate (EPA; 20:5n3)	Lipid	Essential fatty acid	0.015 (0.003)	2.30E-07	0.001 (0.006)	9.09E-01	0.013 (0.003)	1.67E-06
	X-09789			0.021 (0.003)	1.19E-12	0.015 (0.006)	1.00E-02	0.020 (0.003)	2.65E-14
	X-11372			-0.022 (0.003)	7.00E-13	-0.005 (0.006)	3.43E-01	-0.018 (0.003)	6.97E-12
	X-11469			0.019 (0.003)	4.72E-11	0.003 (0.007)	6.77E-01	0.017 (0.003)	2.64E-10
	X-02269			0.019 (0.003)	1.09E-10	0.001 (0.007)	8.54E-01	0.017 (0.003)	7.75E-10
	X-11315			0.015 (0.003)	3.95E-06	0.014 (0.005)	1.42E-02	0.014 (0.003)	1.37E-07
Nuts and legumes	**tryptophan betaine**	Amino acid	Tryptophan metabolism	0.048 (0.007)	9.22E-12	0.017 (0.011)	1.49E-01	0.040 (0.006)	2.52E-11
	X-11315			0.030 (0.006)	7.82E-07	0.023 (0.012)	6.24E-02	0.028 (0.005)	1.18E-07
Seafood	**3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF)**	Lipid	Fatty acid, dicarboxylate	0.149 (0.014)	7.86E-24	0.112 (0.026)	2.84E-05	0.140 (0.012)	4.47E-29
	docosahexaenoate (DHA; 22:6n3)	Lipid	Essential fatty acid	0.167 (0.016)	5.32E-25	0.085 (0.022)	1.96E-04	0.139 (0.013)	1.03E-27
	eicosapentaenoate (EPA; 20:5n3)	Lipid	Essential fatty acid	0.133 (0.015)	4.36E-19	0.100 (0.027)	3.97E-04	0.126 (0.013)	1.25E-22
	1-docosahexaenoylglyc	Lipid	Lysolipid	0.103 (0.013)	4.42E-14	0.050 (0.030)	9.59E-02	0.094 (0.012)	1.47E-14

Table 1. Training group food group metabolite associations

Food group	Metabolite	Super-pathway	Sub-pathway	Discovery		Discordant		Meta-analysis	
				Beta (SE)	*P*	*Beta (SE)*	*P*	*Beta (SE)*	*P*
	ero-phosphocholine*								
	docosapentaenoate (n3 DPA; 22:5n3)	Lipid	Essential fatty acid	0.088 (0.014)	1.62E-10	0.038 (0.023)	1.06E-01	0.075 (0.012)	1.42E-10
	1-arachidonoylglycero-phosphoethanolamine*	Lipid	Lysolipid	-0.068 (0.013)	4.51E-07	-0.117 (0.028)	7.99E-05	-0.077 (0.012)	1.99E-10
	pyroglutamine*	Amino acid	Glutamate metabolism	-0.072 (0.014)	7.09E-07	-0.079 (0.021)	2.52E-04	-0.074 (0.012)	3.54E-10
	1-oleoylglycero-phosphoethanolamine	Lipid	Lysolipid	-0.076 (0.013)	2.99E-09	-0.052 (0.027)	5.66E-02	-0.072 (0.012)	4.28E-10
	pseudouridine	Nucleotide	Pyrimidine metabolism, uracil containing	-0.071 (0.013)	1.04E-07	-0.051 (0.018)	6.92E-03	-0.064 (0.011)	2.40E-09
	1-eicosatrienoylglycero-phosphocholine*	Lipid	Lysolipid	-0.062 (0.013)	2.28E-06	-0.092 (0.026)	7.74E-04	-0.068 (0.012)	6.45E-09
	1-linoleoylglycero-phosphoethanolamine*	Lipid	Lysolipid	-0.064 (0.014)	4.08E-06	-0.058 (0.028)	4.26E-02	-0.063 (0.012)	4.01E-07
	X-11469			0.140 (0.014)	2.17E-21	0.120 (0.028)	4.19E-05	0.136 (0.013)	2.04E-26
	X-02269			0.139 (0.015)	2.27E-20	0.121 (0.027)	2.49E-05	0.135 (0.013)	1.45E-25
	X-11315			0.078 (0.014)	1.82E-08	0.044 (0.025)	7.47E-02	0.070 (0.012)	5.35E-09
	X-12627			0.076 (0.013)	1.61E-08	0.026 (0.023)	2.53E-01	0.064 (0.012)	3.62E-08
	X-12644			0.087 (0.015)	9.74E-09	0.008 (0.019)	6.84E-01	0.057 (0.012)	1.39E-06
	X-11437			0.076 (0.015)	3.97E-07	0.009 (0.027)	7.53E-01	0.061 (0.013)	3.41E-06
	X-12798			-0.061 (0.013)	6.13E-06	-0.035 (0.035)	3.18E-01	-0.058 (0.013)	4.15E-06
White meat	**3-methylhistidine**	Amino acid	Histidine metabolism	0.138 (0.023)	4.46E-09	0.136 (0.046)	3.97E-03	0.138 (0.021)	3.63E-11
Red, processed meat and eggs	**trans-4-hydroxyproline**	Amino acid	Urea cycle; arginine-, proline-, metabolism	0.039 (0.007)	5.30E-08	0.014 (0.015)	3.57E-01	0.034 (0.006)	8.32E-08
	creatine	Amino acid	Creatine metabolism	0.037 (0.007)	6.84E-07	0.016 (0.014)	2.62E-01	0.032 (0.006)	7.36E-07
	X-11381			0.034 (0.007)	6.10E-07	0.023 (0.013)	8.56E-02	0.032 (0.006)	1.44E-07
	X-09789			-0.033 (0.007)	1.65E-06	-0.012 (0.014)	4.01E-01	-0.029 (0.006)	2.65E-06
Fermented dairy	X-11315			0.034 (0.006)	1.71E-08	0.008 (0.012)	5.02E-01	0.029 (0.005)	6.90E-08
Fried foods	3-phenylpropionate (hydrocinnamate)	Amino acid	Phenylalanine & tyrosine metabolism	-0.037 (0.008)	2.82E-06	-0.027 (0.014)	4.79E-02	-0.035 (0.007)	3.59E-07
	X-11372			0.071 (0.009)	2.09E-15	0.045 (0.010)	3.66E-05	0.060 (0.007)	2.93E-19
	X-11469			-0.035 (0.006)	5.23E-08	-0.009 (0.019)	6.36E-01	-0.032 (0.006)	8.19E-08
	X-02269			-0.035 (0.006)	9.23E-08	-0.008 (0.020)	6.80E-01	-0.032 (0.006)	1.52E-07
	X-11315			-0.042 (0.009)	1.24E-06	-0.022 (0.013)	8.69E-02	-0.036 (0.007)	5.26E-07
Sweets and sweet	docosahexaenoate (DHA; 22:6n3)	Lipid	Essential fatty acid	-0.010 (0.002)	7.56E-09	-0.016 (0.002)	3.97E-09	-0.012 (0.001)	1.07E-17

Table 1. Training group food group metabolite associations

Food group	Metabolite	Super-pathway	Sub-pathway	Discovery		Discordant		Meta-analysis	
				Beta (SE)	*P*	*Beta (SE)*	*P*	*Beta (SE)*	*P*
baked products	eicosapentaenoate (EPA; 20:5n3)	Lipid	Essential fatty acid	-0.009 (0.002)	2.20E-08	-0.013 (0.003)	2.14E-06	-0.010 (0.001)	7.29E-14
	pipecolate	Amino acid	Lysine metabolism	-0.010 (0.002)	1.05E-08	-0.011 (0.004)	4.30E-03	-0.010 (0.002)	9.37E-11
	scyllo-inositol	Lipid	Inositol metabolism	-0.012 (0.002)	7.77E-09	-0.011 (0.005)	3.49E-02	-0.012 (0.002)	5.14E-10
	C-glycosyltryptophan*	Amino acid	Tryptophan metabolism	0.008 (0.002)	3.73E-07	0.008 (0.003)	1.09E-02	0.008 (0.001)	9.24E-09
	glycerate	Carbohydrate	Glycolysis, gluconeogenesis, pyruvate metabolism	-0.011 (0.002)	2.89E-07	-0.006 (0.004)	1.10E-01	-0.009 (0.002)	1.13E-07
	1-docosahexaenoylglycero-phosphocholine*	Lipid	Lysolipid	-0.008 (0.002)	3.37E-07	-0.005 (0.004)	1.89E-01	-0.008 (0.001)	1.70E-07
	pyroglutamine*	Amino acid	Glutamate metabolism	0.008 (0.002)	3.05E-06	0.007 (0.004)	9.02E-02	0.007 (0.001)	5.91E-07
	piperine	Xenobiotics	Food component, Plant	-0.009 (0.002)	4.26E-06	-0.007 (0.004)	6.64E-02	-0.008 (0.002)	7.19E-07
	X-11315			-0.013 (0.002)	2.16E-13	-0.004 (0.003)	1.61E-01	-0.011 (0.001)	1.99E-12
	X-11437			-0.010 (0.002)	3.42E-08	-0.002 (0.006)	7.89E-01	-0.009 (0.002)	6.75E-08
	X-12696			0.008 (0.002)	1.84E-06	0.003 (0.004)	4.78E-01	0.007 (0.002)	3.48E-06
Butter and cream	**X-13431--nonanoylcarnitine***	Lipid	Carnitine metabolism	0.034 (0.005)	1.61E-12	0.028 (0.009)	2.67E-03	0.033 (0.004)	6.30E-15
	myristate (14:0)	Lipid	Long chain fatty acid	0.024 (0.005)	1.04E-06	0.015 (0.011)	1.99E-01	0.022 (0.004)	4.99E-07
	caprate (10:0)	Lipid	Medium chain fatty acid	0.020 (0.004)	4.99E-06	0.017 (0.010)	8.37E-02	0.020 (0.004)	9.38E-07
	X-02249			0.028 (0.005)	1.98E-09	0.025 (0.009)	5.53E-03	0.027 (0.004)	2.22E-11
	X-10510			0.020 (0.004)	3.28E-06	0.022 (0.008)	7.81E-03	0.021 (0.004)	6.04E-08
Spreads and dressings	**X-11478**			0.020 (0.003)	1.69E-08	0.012 (0.005)	1.40E-02	0.017 (0.003)	1.22E-09
	X-11521			0.017 (0.003)	8.82E-08	0.013 (0.005)	5.17E-03	0.015 (0.003)	1.22E-09
	X-11261			0.020 (0.004)	1.16E-07	0.015 (0.006)	1.15E-02	0.019 (0.003)	3.94E-09
Milk	alpha-hydroxyisovalerate	Amino acid	Valine, leucine and isoleucine metabolism	-0.055 (0.012)	3.85E-06	-0.048 (0.026)	6.36E-02	-0.054 (0.011)	5.48E-07
	X-21365 [trimethyl-N-aminovalerate]			0.107 (0.012)	6.14E-18	0.076 (0.025)	3.08E-03	0.101 (0.011)	2.04E-20
	X-12798			0.093 (0.012)	3.21E-14	0.047 (0.026)	6.94E-02	0.084 (0.011)	7.72E-15
	X-11381			0.076 (0.011)	2.58E-11	0.024 (0.026)	3.54E-01	0.068 (0.010)	5.01E-11
	X-11795			-0.060 (0.012)	3.58E-07	-0.005 (0.021)	8.03E-01	-0.048 (0.010)	3.93E-06
Soy and other milks	glycerate	Carbohydrate	Glycolysis, gluconeogenesis, pyruvate metabolism	0.150 (0.031)	1.14E-06	0.238 (0.048)	2.75E-04	0.175 (0.026)	1.24E-11
Soda	X-11469			-0.020 (0.004)	1.85E-06	-0.001 (0.012)	9.58E-01	-0.018 (0.004)	5.32E-06
Tea	quinate	Xenobiotics	Food component, Plant	-0.019 (0.002)	6.87E-18	-0.025 (0.005)	1.08E-06	-0.020 (0.002)	1.95E-24
	N-acetylornithine	Amino acid	Urea cycle; arginine-, proline-, metabolism	0.009 (0.002)	2.55E-06	0.008 (0.004)	5.49E-02	0.009 (0.002)	3.14E-07
	pseudouridine	Nucleotide	Pyrimidine	0.008 (0.002)	4.01E-06	0.003 (0.005)	5.02E-01	0.008 (0.002)	4.53E-06

Table 1. Training group food group metabolite associations

Food group	Metabolite	Super-pathway	Sub-pathway	Discovery		Discordant		Meta-analysis	
				Beta (SE)	*P*	*Beta (SE)*	*P*	*Beta (SE)*	*P*
			metabolism, uracil containing						
	X-14473			-0.026 (0.002)	6.70E-34	-0.021 (0.005)	1.29E-04	-0.025 (0.002)	2.82E-40
	X-12734			0.019 (0.002)	1.32E-20	0.002 (0.006)	7.34E-01	0.018 (0.002)	2.90E-20
	X-09789			0.014 (0.002)	3.40E-13	0.006 (0.005)	2.05E-01	0.013 (0.002)	2.01E-13
	X-14374			-0.012 (0.002)	2.59E-10	-0.008 (0.006)	1.88E-01	-0.012 (0.002)	9.06E-11
Coffee	quinate	Xenobiotics	Food component, Plant	0.035 (0.002)	1.08E-43	0.027 (0.005)	4.30E-08	0.033 (0.002)	1.77E-56
	catechol sulfate	Xenobiotics	Benzoate metabolism	0.017 (0.002)	1.43E-12	0.013 (0.005)	1.84E-02	0.016 (0.002)	4.57E-14
	paraxanthine	Xenobiotics	Xanthine metabolism	0.013 (0.003)	8.63E-07	0.008 (0.005)	1.11E-01	0.012 (0.002)	2.77E-07
	theophylline	Xenobiotics	Xanthine metabolism	0.012 (0.003)	3.88E-06	0.007 (0.005)	1.09E-01	0.011 (0.002)	1.26E-06
	X-14473			0.040 (0.002)	1.53E-77	0.028 (0.005)	6.27E-08	0.039 (0.002)	6.41E-102
	X-14374			0.024 (0.002)	1.98E-28	0.021 (0.006)	1.84E-04	0.023 (0.002)	1.08E-33
	X-05426			0.018 (0.002)	4.64E-13	0.008 (0.005)	9.86E-02	0.016 (0.002)	2.01E-13
	X-12217			0.014 (0.003)	1.64E-07	0.010 (0.006)	1.03E-01	0.013 (0.002)	3.82E-08
	X-09789			-0.014 (0.002)	6.43E-08	-0.004 (0.005)	4.04E-01	-0.012 (0.002)	1.43E-07
	X-12734			-0.014 (0.003)	1.20E-06	-0.007 (0.006)	2.41E-01	-0.013 (0.003)	8.02E-07
Alcohol	scyllo-inositol	Lipid	Inositol metabolism	0.042 (0.005)	2.16E-18	0.037 (0.011)	1.83E-03	0.041 (0.004)	1.57E-21
	alpha-hydroxyisovalerate	Amino acid	Valine, leucine and isoleucine metabolism	0.035 (0.005)	1.29E-12	0.034 (0.009)	4.11E-04	0.035 (0.004)	3.98E-16
	4-androsten-3beta,17beta-diol disulfate 1*	Lipid	Sterol, Steroid	0.032 (0.004)	1.25E-12	0.037 (0.011)	1.64E-03	0.032 (0.004)	1.89E-15
	eicosapentaenoate (EPA; 20:5n3)	Lipid	Essential fatty acid	0.019 (0.004)	2.26E-07	0.024 (0.005)	7.77E-05	0.021 (0.003)	1.57E-11
	docosahexaenoate (DHA; 22:6n3)	Lipid	Essential fatty acid	0.020 (0.004)	4.03E-08	0.024 (0.007)	1.57E-03	0.020 (0.003)	1.09E-10
	2-oleoylglycero-phosphocholine*	Lipid	Lysolipid	0.016 (0.003)	2.32E-07	0.030 (0.008)	6.93E-04	0.018 (0.003)	6.90E-10
	docosapentaenoate (n3 DPA; 22:5n3)	Lipid	Essential fatty acid	0.019 (0.004)	2.58E-07	0.025 (0.009)	1.15E-02	0.020 (0.003)	7.10E-09
	pipecolate	Amino acid	Lysine metabolism	0.020 (0.004)	6.11E-06	0.021 (0.008)	1.31E-02	0.020 (0.004)	1.71E-07
	1-docosahexaenoylglycero-phosphocholine*	Lipid	Lysolipid	0.015 (0.003)	1.39E-06	0.012 (0.009)	1.65E-01	0.015 (0.003)	4.40E-07
	X-11795			0.030 (0.004)	1.11E-11	0.036 (0.009)	1.36E-04	0.031 (0.004)	1.15E-15
	X-10395			0.026 (0.004)	6.35E-11	0.019 (0.010)	6.55E-02	0.025 (0.004)	7.70E-12
	X-14473			0.020 (0.003)	1.06E-09	0.023 (0.009)	1.06E-02	0.020 (0.003)	1.98E-11
	X-12627			0.019 (0.003)	2.06E-08	0.016 (0.008)	5.24E-02	0.018 (0.003)	2.13E-09
	X-11204			0.017 (0.003)	2.34E-07	0.015 (0.008)	5.40E-02	0.017 (0.003)	2.61E-08
	X-11444			-0.018 (0.003)	9.08E-08	-0.012 (0.008)	1.43E-01	-0.017 (0.003)	3.13E-08
	X-12644			0.014 (0.003)	6.55E-06	0.017 (0.008)	2.86E-02	0.014 (0.003)	4.46E-07
	X-11787			-0.014 (0.003)	2.67E-06	-0.007 (0.009)	4.12E-01	-0.014 (0.003)	2.22E-06

Notes: Table shows results of the linear regression analysis for the discovery population (excluding monozygotic twins discordant for each food group), the MZ discordant twin sample and the fixed effects meta-analysis of both groups. Only significant associations are shown which includes those associations passing the bonferroni cut-off in the discovery and fixed effects analyses (7.60×10^{-6} = 0.05/[20 food groups x 329 detected metabolites]) and passing the 5% level of significance in the discordant twin group. The Metabolon platform is a non-targeted platform which identified 456 metabolites in blood for which data were available for 1780 twins.

Table 2. Multivariate regression results for metabolites associated with multiple food groups

Superpathway	Metabolite	Food group	beta (SE)	*P*
Amino acid	3-phenylpropionate (hydrocinnamate)	Fried foods	-0.028 (0.007)	2.44E-05
		Whole grains	0.012 (0.003)	6.98E-05
	alpha-hydroxyisovalerate	**Alcohol**	**0.033 (0.004)**	**1.37E-14***
		Milk	-0.038 (0.011)	3.60E-04
	pipecolate	**Sweets and sweet baked products**	**-0.008 (0.002)**	**1.78E-06***
		Alcohol	0.016 (0.004)	2.77E-05
	pyroglutamine*	**Seafood**	**-0.064 (0.012)**	**2.31E-07***
		Sweets and sweet baked products	0.006 (0.002)	2.35E-04
Carbohydrate	**glycerate**	**Fruit**	**0.013 (0.002)**	**2.38E-11***
		Soy and other milk	0.132 (0.030)	9.40E-06
		Sweets and sweet baked products	-0.006 (0.002)	1.97E-04
Lipid	**1-docosahexaenoylglycerophosphocholine***	**Seafood**	**0.083 (0.012)**	**3.08E-12***
		Alcohol	0.010 (0.003)	1.09E-03
		Sweets and sweet baked products	-0.004 (0.002)	7.12E-03
	docosahexaenoate (DHA; 22:6n3)	**Seafood**	**0.127 (0.012)**	**1.79E-24***
		Alcohol	**0.015 (0.003)**	**7.79E-08***
		Sweets and sweet baked products	-0.005 (0.002)	1.71E-03
		Fruit	0.005 (0.002)	8.25E-03
	docosapentaenoate (n3 DPA; 22:5n3)	**Seafood**	**0.070 (0.011)**	**2.94E-10***
		Alcohol	**0.017 (0.003)**	**1.39E-07***
	eicosapentaenoate (EPA; 20:5n3)	**Seafood**	**0.112 (0.012)**	**2.24E-19***
		Alcohol	**0.015 (0.003)**	**7.21E-08***
		Whole grains	0.010 (0.003)	7.59E-05
		Sweets and sweet baked products	-0.004 (0.001)	1.96E-03
	scyllo-inositol	**Alcohol**	**0.042 (0.004)**	**3.03E-22***
		Fruit	**0.013 (0.002)**	**7.22E-10***
		Sweets and sweet baked products	-0.004 (0.002)	3.16E-02
Nucleotide	**pseudouridine**	**Seafood**	**-0.065 (0.011)**	**1.76E-09***
		Tea	0.007 (0.002)	1.85E-05
Xenobiotics	**quinate**	**Coffee**	**0.029 (0.002)**	**4.90E-35***
		Tea	**-0.012 (0.002)**	**1.66E-09***
Unknown	**X-02269**	**Seafood**	**0.127 (0.012)**	**6.27E-24***
		Whole grains	**0.012 (0.003)**	**3.52E-06***
		Fried foods	-0.014 (0.007)	3.75E-02
	X-09789	**Whole grains**	**0.017 (0.003)**	**3.95E-11***
		Tea	**0.011 (0.002)**	**9.24E-10***
		Red, processed meat and eggs	-0.022 (0.006)	3.20E-04
		Coffee	-0.007 (0.002)	2.48E-03
	X-11315	**Fruit**	**0.014 (0.002)**	**1.21E-10***
		Fermented dairy	0.017 (0.005)	1.17E-03
		Seafood	0.037 (0.012)	2.03E-03
		Sweets and sweet baked products	-0.005 (0.002)	2.61E-03
		Whole grains	0.007 (0.003)	1.23E-02
		Nuts and legumes	0.012 (0.006)	3.39E-02
		Fried foods		5.16E-02
		Vegetable		2.60E-01
	X-11372	**Fried foods**	**0.052 (0.007)**	**9.20E-15***

Table 2. Multivariate regression results for metabolites associated with multiple food groups

Superpathway	Metabolite	Food group	beta (SE)	*P*
		Whole grains	-0.012 (0.003)	1.03E-05
		Fruit	-0.004 (0.002)	4.39E-02
		Vegetable	-0.003 (0.001)	4.85E-02
	X-11381	**Milk**	**0.070 (0.010)**	**9.83E-12***
		Red, processed meat and eggs	**0.035 (0.006)**	**1.39E-08***
	X-11437	Seafood	0.047 (0.013)	5.05E-04
		Sweets and sweet baked products	-0.006 (0.002)	6.14E-04
	X-11469	**Seafood**	**0.127 (0.012)**	**2.13E-24***
		Whole grains	**0.012 (0.003)**	**2.89E-06***
		Soda	-0.012 (0.004)	1.74E-03
		Fried foods	-0.014 (0.007)	4.25E-02
	X-11795	**Alcohol**	**0.029 (0.004)**	**1.48E-13***
		Milk	-0.035 (0.010)	6.93E-04
	X-12627	**Alcohol**	**0.017 (0.003)**	**9.29E-09***
		Seafood	**0.057 (0.011)**	**1.70E-07***
	X-12644	**Seafood**	**0.067 (0.012)**	**2.25E-08***
		Alcohol	0.012 (0.003)	3.48E-05
	X-12734	**Tea**	**0.016 (0.002)**	**2.55E-14***
		Coffee	-0.006 (0.003)	2.38E-02
	X-12798	**Milk**	**0.080 (0.011)**	**8.80E-13***
		Seafood	-0.042 (0.013)	1.43E-03
	X-14374	**Coffee**	**0.020 (0.002)**	**2.54E-22***
		Tea	-0.007 (0.002)	3.88E-04
	X-14473	**Coffee**	**0.032 (0.002)**	**4.05E-57***
		Tea	**-0.016 (0.002)**	**1.47E-17***
		Alcohol	0.011 (0.003)	1.41E-04

Notes: Table shows results of a multivariate backwards stepwise linear regression including all food groups associated to a metabolite from the discovery analysis using *P*<0.05 as the cut-off threshold. Metabolites and foods in bold, and with an asterisk (*) beside the *P*-value passed the Bonferroni cut-off for multiple testing from the discovery analysis (7.60×10^{-6}) and were considered significant. The analysis was conducted in 1780 twins.

Table 3. Backward stepwise linear regression results for foods associated with multiple metabolites and foods associated with single metabolites

		Model 1			Model 2			Standardized betas
Food group	**Metabolite**	R^2	Beta(SE)	P	R^2	Beta(SE)	P	
Vegetables	X-12063	0.0094	-1.798 (0.627)	4.23E-03				-0.101
Fruit	stachydrine	0.1195	2.327 (0.345)	2.59E-11				0.194
	X-11315		2.586 (0.391)	6.16E-11				0.207
	proline		-1.506 (0.346)	1.48E-05				-0.119
	threitol		1.614 (0.409)	8.63E-05				0.097
	hippurate			5.46E-02				
	glycerate			6.20E-02				
	scyllo-inositol			0.454				
	threonate			0.798				
Whole grains	X-09789	0.0508	1.552 (0.229)	2.08E-11				0.173
	X-11469		1.337 (0.235)	1.79E-08				0.142
	X-02269			0.650				0.173
Nuts and legumes	tryptophan betaine	0.0384	0.905 (0.132)	1.32E-11				0.181
Seafood	docosahexaenoate (DHA; 22:6n3)	0.1514	0.563 (0.090)	7.03E-10	0.1535	0.399 (0.056)	2.09E-12	0.192
	X-11469		0.246 (0.063)	9.44E-05		0.281 (0.061)	3.89E-06	0.130
	pseudouridine		-0.165 (0.047)	5.13E-04		-0.164 (0.046)	4.31E-04	-0.079
	1-eicosatrienoylglycerophosphocholine*		-0.182 (0.067)	6.71E-03		-0.205 (0.063)	1.29E-03	-0.105
	pyroglutamine*		-0.133 (0.050)	8.34E-03		-0.130 (0.049)	8.44E-03	-0.065
	docosapentaenoate (n3 DPA; 22:5n3)		-0.194 (0.076)	1.08E-02				
	1-arachidonoylglycerophosphoethanolamine*		-0.110 (0.053)	3.99E-02		-0.123 (0.051)	1.52E-02	-0.062
	1-docosahexaenoylglycerophosphocholine*		0.162 (0.081)	4.52E-02		0.169 (0.078)	3.01E-02	0.084
	eicosapentaenoate (EPA; 20:5n3)			6.37E-02				
	1-linoleoylglycerophosphoethanolamine*			7.63E-02				
	1-oleoylglycerophosphoethanolamine			8.13E-02				
	X-02269			0.255				
	3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF)			0.271				
	X-12627			0.450				
	X-12644			0.580				
White meat	3-methylhistidine	0.0291	0.219 (0.034)	1.62E-10				0.170
Red meat	trans-4-hydroxyproline	0.0403	0.449 (0.101)	9.20E-06				0.113
	X-11381		0.393 (0.093)	2.83E-05				0.102
	creatine		0.401 (0.099)	5.40E-05				0.102
Fried foods	X-11372	0.0668	1.058 (0.100)	9.39E-25				0.279
Sweets and sweet baked	C-glycosyltryptophan*	0.0588	2.233 (0.500)	8.93E-06				0.129

Table 3. Backward stepwise linear regression results for foods associated with multiple metabolites and foods associated with single metabolites

		Model 1			Model 2			Standardized betas
Food group	**Metabolite**	R^2	*Beta(SE)*	*P*	R^2	*Beta(SE)*	*P*	
products	pipecolate		-1.719 (0.439)	9.77E-05				-0.117
	X-12696		1.582 (0.435)	2.94E-04				0.102
	piperine		-1.406 (0.405)	5.40E-04				-0.093
Butter and cream	nonanoylcarnitine*	0.0754	1.030 (0.188)	5.95E-08				0.164
	X-02249		0.615 (0.185)	8.99E-04				0.093
	myristate (14:0)		0.511 (0.175)	3.52E-03				0.083
	caprate (10:0)		0.363 (0.168)	3.05E-02				0.058
	X-10510		0.349 (0.161)	3.09E-02				0.053
Spreads and dressings	X-11261	0.0331	1.103 (0.355)	1.94E-03				0.101
	X-11478		0.774 (0.291)	8.00E-03				0.093
Milk	X-21365 [trimethyl-N-aminovalerate]	0.0735	0.401 (0.065)	1.31E-09				0.175
	X-11381		0.206 (0.059)	4.88E-04				0.091
	X-12798		0.238 (0.069)	6.52E-04				0.102
Tea	X-14473	0.2113	-3.959 (0.503)	1.33E-14				-0.282
	N-acetylornithine		1.710 (0.376)	6.26E-06				0.122
	X-12734		1.675 (0.390)	2.01E-05				0.121
	X-09789		1.461 (0.430)	7.15E-04				0.102
	quinate		-1.253 (0.495)	1.16E-02				-0.090
Coffee	X-14473	0.1963	3.756 (0.404)	2.02E-19				0.344
	quinate		2.197 (0.439)	7.27E-07				0.204
	X-14374			5.46E-02				
	X-05426			0.115				
	catechol sulfate			0.313				
	X-12217			0.715				
	theophylline			0.887				
	paraxanthine			0.964				
Alcohol	scyllo-inositol	0.2537	1.813 (0.213)	1.04E-16	0.2493	1.776 (0.209)	1.16E-16	0.231
	4-androsten-3beta,17beta-diol disulfate 1*		1.692 (0.272)	8.18E-10		1.623 (0.271)	3.21E-09	0.188
	X-11795		1.107 (0.209)	1.57E-07		1.141 (0.206)	4.39E-08	0.143
	X-11444		-1.130 (0.224)	5.74E-07		-1.075 (0.221)	1.33E-06	-0.131
	alpha-hydroxyisovalerate		1.066 (0.224)	2.28E-06		1.010 (0.213)	2.54E-06	0.127
	eicosapentaenoate (EPA; 20:5n3)		1.042 (0.281)	2.23E-04		0.827 (0.222)	2.06E-04	0.098
	2-oleoylglycerophosphocholine*		0.630 (0.206)	2.30E-03		0.626 (0.203)	2.11E-03	0.080
	X-11787		-0.491 (0.217)	2.42E-02		-0.442 (0.217)	4.18E-02	-0.053
	X-12627		-0.484 (0.241)	4.51E-02				
	X-10395			0.112				
	docosapentaenoate (n3 DPA; 22:5n3)			0.489				
	docosahexaenoate (DHA; 22:6n3)			0.502				
	X-11204			0.928				

Notes: Table shows the results of a multivariate backwards stepwise linear regression including all associated metabolites used to predict food group intakes using *P*<0.05 as the cut-off threshold. Metabolites highlighted in red indicate the association was in the opposite direction to the discovery analysis for model 1 and therefore, for the corresponding food group a second multivariate regression was undertaken. The rightmost column shows the final standardized betas used for the weighted scoring method. The analysis was conducted in 1780 twins.

Table 4. Associations between reported food intakes and the final scores in the training and test groups

			Training group		Test group	
Food group	**No. metabolites in score**	**Score**	**Beta (SE)**	***P***	**Beta (SE)**	***P***
Vegetables	1	Top: X-12063	-0.005 (0.001)	2.96E-06	-0.006 (0.002)	3.62E-04*
		Quartiles	0.004 (0.001)	8.59E-03	0.005 (0.002)	6.88E-03*
Fruit	4	Top: X-11315	0.019 (0.002)	7.14E-20	0.016 (0.002)	1.81E-18
		Quartiles	0.064 (0.005)	4.06E-34	0.062 (0.005)	1.28E-36
		Continuous	0.060 (0.005)	2.68E-36	0.055 (0.004)	3.76E-37
		Weighted	0.010 (0.001)	3.03E-36	0.009 (0.001)	4.48E-40
Whole grains	3	Top: X-09789	0.020 (0.003)	1.33E-13	0.031 (0.003)	4.35E-22
		Quartiles	0.038 (0.005)	6.99E-16	0.049 (0.005)	5.42E-23
		Continuous	0.036 (0.004)	1.40E-18	0.046 (0.005)	6.98E-22
		Weighted	0.006 (0.001)	8.78E-19	0.008 (0.001)	9.51E-23
Nuts and legumes	1	Top: tryptophan betaine	0.044 (0.006)	1.45E-13	0.037 (0.005)	3.38E-13
		Quartiles	0.041 (0.005)	2.78E-14	0.037 (0.005)	1.29E-13
Seafood	7	Top: DHA	0.147 (0.013)	6.98E-30	0.141 (0.011)	1.76E-34
		Quartiles	0.711 (0.048)	1.24E-45	0.599 (0.049)	7.85E-32
		Continuous	0.662 (0.046)	1.30E-42	0.575 (0.045)	3.43E-34
		Weighted	0.076 (0.005)	6.01E-44	0.069 (0.005)	1.63E-41
White meat	1	Top: 3-methylhistidine	0.135 (0.021)	1.52E-10	0.077 (0.023)	7.78E-04
		Quartiles	0.176 (0.021)	2.11E-16	0.134 (0.022)	1.03E-09
Red meat	3	Top: trans-4-hydroxyproline	0.035 (0.006)	9.27E-08	0.036 (0.007)	3.66E-08
		Quartiles	0.108 (0.013)	9.64E-16	0.124 (0.014)	9.23E-19
		Continuous	0.100 (0.012)	2.96E-16	0.114 (0.012)	3.43E-19
		Weighted	0.011 (0.001)	3.40E-16	0.012 (0.001)	3.79E-19
Fried foods	1	Top: X-11372	0.067 (0.007)	3.28E-20	0.080 (0.009)	2.55E-17
		Quartiles	0.068 (0.008)	2.30E-18	0.084 (0.010)	1.31E-16
Sweets and sweet baked products	4	Top: C-glycosyltryptophan*	0.136 (0.034)	8.54E-05	0.008 (0.001)	1.78E-09
		Quartiles	0.457 (0.095)	2.00E-06	0.032 (0.004)	6.44E-17
		Continuous	0.427 (0.082)	2.27E-07	0.031 (0.003)	8.93E-21
		Weighted	0.047 (0.009)	1.73E-07	0.003 (0.000)	3.28E-21
Butter and creams	5	Top: nonanoylcarnitine*	0.034 (0.004)	2.69E-16	0.022 (0.004)	8.92E-08
		Quartiles	0.122 (0.013)	2.53E-21	0.107 (0.013)	4.89E-16
		Continuous	0.120 (0.013)	1.90E-18	0.094 (0.012)	8.66E-15
		Weighted	0.011 (0.001)	8.09E-21	0.009 (0.001)	2.30E-15
Spreads and dressings	2	Top: X-11261	0.019 (0.003)	2.77E-09	0.014 (0.003)	6.34E-06
		Quartiles	0.041 (0.006)	4.21E-12	0.034 (0.007)	4.86E-07
		Continuous	0.036 (0.005)	5.47E-11	0.026 (0.006)	1.83E-05*
		Weighted	0.003 (0.001)	5.89E-11	0.002 (0.001)	1.68E-05*
Milk	3	Top: X-21365	0.102 (0.011)	2.58E-20	0.094 (0.010)	6.20E-19
		Quartiles	0.264 (0.026)	3.02E-23	0.247 (0.028)	1.60E-18
		Continuous	0.250 (0.023)	1.23E-25	0.204 (0.024)	4.95E-17
		Weighted	0.032 (0.003)	2.90E-26	0.027 (0.003)	1.49E-19
Tea	5	Top: X-14473	-0.025 (0.002)	2.56E-37	-0.024 (0.002)	2.83E-36
		Quartiles	0.078 (0.005)	3.98E-42	0.050 (0.005)	3.32E-22
		Continuous	0.073 (0.005)	2.96E-50	0.051 (0.004)	2.08E-32
		Weighted	0.012 (0.001)	1.15E-52	0.009 (0.001)	7.64E-38

Table 4. Associations between reported food intakes and the final scores in the training and test groups

			Training group		Test group	
Food group	**No. metabolites in score**	**Score**	**Beta (SE)**	***P***	**Beta (SE)**	***P***
Coffee	2	Top: X-14473	0.038 (0.002)	3.47E-83	0.036 (0.002)	3.00E-61
		Quartiles	0.090 (0.003)	8.56E-115	0.091 (0.004)	1.06E-104
		Continuous	0.064 (0.003)	2.63E-82	0.057 (0.003)	1.66E-71
		Weighted	0.018 (0.001)	4.83E-87	0.016 (0.001)	5.92E-72
Alcohol	8	Top: scyllo-inositol	0.041 (0.004)	8.22E-22	0.039 (0.004)	2.48E-24
		Quartiles	0.212 (0.018)	7.53E-31	0.191 (0.013)	1.21E-42
		Continuous	0.197 (0.018)	1.06E-27	0.180 (0.013)	2.51E-41
		Weighted	0.029 (0.003)	5.47E-27	0.027 (0.002)	7.79E-45

Notes: A linear regression was performed in the training (*n*=1780) and test groups (*n*=1779) using each food group intake as a predictor of the top metabolite and metabolite scores. An asterisk (*) beside the *P*-value indicates the association does not pass the Bonferroni cut-off from the discovery analysis ($P<7.60\text{x}10^{-6}$)

Table 5. Average (SD) energy-adjusted intakes (servings/week) for the top and bottom tertiles for each food in the test group

Food group	Tertile 1 (*n*=593)	Tertile 3 (*n*=593)
Vegetables	19.5 (5.4)	51.0 (13.4)
Fruit	9.7 (4.4)	35.6 (10.3)
Whole grains	2.7 (2.4)	19.2 (7.4)
Nuts and legumes	3.0 (1.5)	12.9 (5.4)
Seafood	0.7 (0.5)	4.5 (1.8)
White meat	0.6 (0.4)	3.4 (0.8)
Red meat	2.9 (1.5)	10.8 (3.0)
Fried foods	1.6 (1.1)	8.0 (3.3)
Sweets and sweet baked products	4.2 (3.6)	30.4 (14.5)
Butter and cream	-0.4 (1.0)	10.6 (7.4)
Spreads and dressings	1.7 (2.1)	17.3 (10.8)
Milk	1.1 (0.9)	5.8 (1.5)
Tea	3.2 (3.7)	35.5 (5.6)
Coffee	0.0 (0.5)	22.1 (7.9)
Alcohol	0.4 (1.0)	14.3 (8.8)

Table 6. Results of ROC analysis

						Versus top metabolite	
Food group	**Score**	**Sensitivity**	**Specificity**	**Correctly classified**	**ROC area**	χ^2	***P***
Vegetables	Top: X-12063	59.74%	55.47%	57.60%	0.6136 [0.5802, 0.6470]		
	Quartiles	59.22%	54.50%	56.85%	0.6136 [0.5802, 0.6470]	0.00	1.0000
Fruit	Top: X-11315	65.59%	60.34%	62.98%	0.6783 [0.6479, 0.7087]		
	Quartiles	68.54%	67.06%	67.80%	0.7305 [0.7020, 0.7590]	18.14	<0.0001
	Continuous	67.52%	66.55%	67.04%	0.7339 [0.7057, 0.7622]	19.99	<0.0001
	Weighted	**66.84%**	**67.24%**	**67.04%**	**0.7411 [0.7131, 0.7690]**	**29.77**	**<0.0001**
Whole grains	Top: X-09789	65.25%	67.07%	66.15%	0.7134 [0.6841, 0.7428]		
	Quartiles	66.61%	65.47%	66.04%	0.7063 [0.6768, 0.7359]	0.60	0.4394
	Continuous	67.46%	66.32%	66.89%	0.7186 [0.6895, 0.7478]	0.36	0.5500
	Weighted	67.63%	65.81%	66.72%	0.7224 [0.6934, 0.7514]	1.37	0.2425
Nuts and legumes	Top: tryptophan betaine	64.25%	60.70%	62.47%	0.6723 [0.6366, 0.7080]		
	Quartiles	65.82%	57.68%	61.77%	0.6629 [0.6268, 0.6991]	2.42	0.1197
Seafood	Top: DHA	64.79%	67.46%	66.13%	0.7230 [0.6943, 0.7516]		
	Quartiles	65.81%	67.12%	66.47%	0.7307 [0.7023, 0.7592]	0.42	0.5172
	Continuous	66.15%	67.12%	66.64%	0.7356 [0.7075, 0.7637]	1.12	0.2893
	Weighted	**69.57%**	**70.68%**	**70.13%**	**0.7547 [0.7274, 0.7821]**	**11.88**	**0.0006**
White meat	**Top: 3-methylhistidine**	**77.78%**	**32.44%**	**56.86%**	**0.5733 [0.5371, 0.6096]**		
	Quartiles	54.31%	62.39%	58.40%	0.5536 [0.5172, 0.5901]	4.74	0.0295
Red meat	Top: trans-4-hydroxyproline	58.94%	65.37%	62.16%	0.6697 [0.6390, 0.7003]		
	Quartiles	61.73%	63.50%	62.62%	0.6920 [0.6622, 0.7218]	5.36	0.0207
	Continuous	63.10%	65.70%	64.40%	0.6963 [0.6666, 0.7260]	8.63	0.0033
	Weighted	**63.61%**	**65.53%**	**64.57%**	**0.6963 [0.6666, 0.7260]**	**9.36**	**0.0022**
Fried foods	**Top:** X-11372	**65.30%**	**66.32%**	**65.81%**	**0.7322 [0.7038, 0.7606]**		
	Quartiles	65.98%	65.93%	65.96%	0.7217 [0.6928, 0.7506]	6.70	0.0096
Sweets and sweet baked products	Top: C-glycosyltryptophan*	58.67%	56.51%	57.59%	0.6002 [0.5679, 0.6325]		
	Quartiles	58.57%	57.78%	58.18%	0.6235 [0.5917, 0.6552]	3.29	0.0697
	Continuous	60.27%	59.32%	59.80%	0.6249 [0.5932, 0.6566]	4.29	0.0384
	Weighted	**60.10%**	**58.97%**	**59.54%**	**0.6254 [0.5937, 0.6571]**	**5.07**	**0.0243**
Butter and cream	Top: nonanoylcarnitine*	71.88%	39.55%	56.84%	0.5889 [0.5529, 0.6249]		
	Quartiles	61.02%	58.87%	59.95%	0.6386 [0.6037, 0.6735]	7.45	0.0064
	Continuous	58.31%	60.75%	59.52%	0.6383 [0.6034, 0.6732]	7.38	0.0066
	Weighted	**59.15%**	**59.90%**	**59.52%**	**0.6385 [0.6036, 0.6734]**	**11.39**	**0.0007**
Spreads and dressings	Top: X-11261	59.08%	57.48%	58.28%	0.6199 [0.5880, 0.6518]		
	Quartiles	**60.27%**	**60.10%**	**60.19%**	**0.6374 [0.6059, 0.6689]**	**5.42**	**0.0199**
	Continuous	59.25%	60.61%	59.93%	0.6319 [0.6003, 0.6636]	3.07	0.0799
	Weighted	58.90%	60.27%	59.59%	0.6316 [0.5999, 0.6633]	3.14	0.0762
Milk	Top: X-21365	60.17%	60.65%	60.41%	0.6409 [0.6096, 0.6723]		
	quartiles	60.51%	61.33%	60.92%	0.6403 [0.6089, 0.6717]	0.00	0.9575
	Continuous	60.68%	61.33%	61.00%	0.6355 [0.6040, 0.6669]	0.22	0.6402
	Weighted	60.68%	60.48%	60.58%	0.6442 [0.6130, 0.6755]	0.14	0.7122
Tea	Top: X-14473	59.28%	73.51%	67.04%	0.7384 [0.7078, 0.7691]		
	Quartiles	61.63%	64.63%	63.13%	0.7121 [0.6803, 0.7439]	2.68	0.1016
	Continuous	66.04%	66.16%	66.10%	0.7257 [0.6943, 0.7572]	0.80	0.3705
	Weighted	67.40%	65.65%	66.53%	0.7443 [0.7139, 0.7747]	0.40	0.5276

Table 6. Results of ROC analysis

						Versus top metabolite	
Food group	**Score**	**Sensitivity**	**Specificity**	**Correctly classified**	**ROC area**	***χ²***	***P***
Coffee	Top: X-14473	78.21%	65.38%	72.42%	0.7779 [0.7483, 0.8074]		
	Quartiles	72.06%	76.36%	74.21%	0.7917 [0.7627, 0.8208]	3.08	0.0790
	Continuous	66.78%	70.92%	68.85%	0.7803 [0.7510, 0.8096]	0.11	0.7388
	Weighted	67.63%	70.75%	69.19%	0.7856 [0.7565, 0.8148]	2.06	0.1511
Alcohol	Top: scyllo-inositol	75.75%	63.32%	70.02%	0.7531 [0.7230, 0.7832]		
	Quartiles	71.48%	70.97%	71.22%	0.7914 [0.7638, 0.8191]	5.52	0.0188
	Continuous	70.80%	74.02%	72.41%	0.8067 [0.7800, 0.8333]	11.05	0.0009
	Weighted	**74.70%**	**76.40%**	**75.55%**	**0.8351 [0.8104, 0.8598]**	**38.99**	**<0.0001**

Notes: Table shows the results of the receiver operating characteristic analysis testing the ability of each top metabolite or metabolite score to predict the top (1, positive outcome) and bottom (0, negative outcome) tertiles of food group intake. The area under the receiver operating characteristic curve (AUC) was compared for each metabolite score against the top performing metabolite. Metabolite scores which performed the best and better than the top metabolite are in bold. The analysis was conducted in the test group (*n*=1779).

Table 7. Top metabolite and score associations with food preferences in the test group

			Liking	
Food group	No. metabolites in score	Score	Beta (SE)	*P*
Vegetables	1	Top: X-12063	-0.062 (0.059)	*NS*
		Quartiles	0.036 (0.067)	*NS*
Fruit	4	Top: X-11315	0.203 (0.045)	7.40E-06
		Quartiles	0.490 (0.122)	6.62E-05
		Continuous	0.448 (0.108)	3.51E-05
		Weighted	0.077 (0.017)	1.12E-05
Whole grains	3	Top: X-09789	0.287 (0.054)	1.37E-07
		Quartiles	0.226 (0.083)	6.62E-03*
		Continuous	0.258 (0.075)	6.29E-04
		Weighted	0.045 (0.012)	1.70E-04
Nuts and legumes	1	Top: tryptophan betaine	0.271 (0.045)	3.04E-09
		Quartiles	0.293 (0.042)	9.87E-12
Seafood	7	Top: DHA	0.333 (0.035)	6.85E-20
		Quartiles	1.538 (0.139)	2.12E-26
		Continuous	1.464 (0.129)	1.01E-27
		Weighted	0.173 (0.014)	1.51E-30
White meat	1	Top: 3-methylhistidine	0.083 (0.037)	2.45E-02*
		Quartiles	0.118 (0.035)	8.28E-04
Red meat	3	Top: trans-4-hydroxyproline	0.221 (0.046)	2.29E-06
		Quartiles	0.854 (0.093)	4.01E-19
		Continuous	0.766 (0.083)	3.54E-19
		Weighted	0.081 (0.009)	7.71E-19
Fried foods	1	Top: X-11372	0.344 (0.046)	2.69E-13
		Quartiles	0.355 (0.051)	1.12E-11
Sweets and sweet baked products	4	Top: C-glycosyltryptophan*	0.004 (0.002)	1.09E-02*
		Quartiles	0.021 (0.004)	5.24E-07
		Continuous	0.020 (0.004)	2.07E-07
		Weighted	0.002 (0.000)	2.24E-07
Butter and creams	5	Top: nonanoylcarnitine*	0.023 (0.033)	4.79E-01*
		Quartiles	0.294 (0.094)	1.78E-03
		Continuous	0.262 (0.080)	1.11E-03
		Weighted	0.022 (0.007)	2.61E-03
Spreads and dressings	2	Top: X-11261	0.105 (0.041)	1.15E-02*
		Quartiles	0.207 (0.089)	2.06E-02*
		Continuous	0.203 (0.073)	5.71E-03*
		Weighted	0.020 (0.007)	5.67E-03*
Milk	3	Top: X-21365	0.167 (0.041)	5.71E-05
		Quartiles	0.384 (0.116)	9.55E-04
		Continuous	0.352 (0.097)	3.15E-04
		Weighted	0.047 (0.012)	9.71E-05
Tea	5	Top: X-14473	-0.217 (0.031)	7.83E-12
		Quartiles	0.624 (0.082)	8.88E-14
		Continuous	0.549 (0.068)	3.19E-15
		Weighted	0.089 (0.011)	2.53E-16
Coffee	2	Top: X-14473	0.241 (0.055)	1.36E-05
		Quartiles	0.689 (0.101)	1.98E-11
		Continuous	0.379 (0.073)	2.50E-07
		Weighted	0.104 (0.020)	4.38E-07
Alcohol	8	Top: scyllo-inositol	0.322 (0.042)	1.01E-13
		Quartiles	1.559 (0.144)	1.22E-25
		Continuous	1.299 (0.125)	6.46E-24
		Weighted	0.197 (0.018)	4.08E-27

NS, not significant; *, not statistically significant ($P<3.33x10^{-3}$).

Notes: Table shows the results of the linear regression using each food group liking score used to predict the top metabolite and metabolite score. The analysis was conducted in the test group (*n*=1779).

Appendix E. Chapter 6 Appendices

Table 1. List of OTUs associated with hippurate, the hippurate diet score and foods [1]

OTU ID [2]	Assigned taxonomy	Hippurate		Diet score		Foods [3]
		Beta (SE)	*P*	Beta (SE)	*P*	(*P*<0.05)
denovo467	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Actinomycetaceae; g__Actinomyces; s__	-0.099 (0.022)	5.14E-06	-0.051 (0.011)	2.81E-06	Fruit: -0.005(0.002) WG: -0.009(0.003)
denovo44	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Rikenellaceae; g__; s__	0.094 (0.023)	6.55E-05			
denovo299	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__	0.113 (0.024)	2.21E-06	0.044 (0.010)	9.76E-06	Coffee: 0.013(0.002)*
denovo346	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__	0.129 (0.025)	2.07E-07			
denovo197	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__; s__	0.126 (0.026)	1.44E-06			
denovo52	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__Clostridium; s__	0.117 (0.024)	1.50E-06			
denovo1263	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae	-0.097 (0.022)	1.38E-05			
denovo272	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__	-0.094 (0.022)	1.29E-05			
denovo20	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__	-0.117 (0.022)	1.29E-07			
denovo100	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__	-0.107 (0.022)	1.92E-06			
denovo55	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__[Ruminococcus]; s__	-0.123 (0.021)	1.17E-08	-0.054 (0.011)	2.79E-06	Fruit: -0.006(0.002)* WG: -0.009(0.003)
denovo27	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__[Ruminococcus]; s__gnavus	-0.107 (0.023)	3.04E-06	-0.064 (0.011)	1.99E-08	Fruit: -0.006(0.002)* WG: -0.009(0.003)
denovo25	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Blautia; s__	-0.135 (0.024)	1.28E-08			
denovo35	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Blautia; s__	-0.137 (0.022)	4.94E-10			
denovo13	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Blautia; s__	-0.177 (0.021)	5.56E-17			
denovo144	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Blautia; s__	-0.165 (0.022)	4.83E-14			
denovo237	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Blautia; s__producta	-0.102 (0.022)	4.78E-06			
denovo329	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Dorea; s__	-0.130 (0.023)	1.73E-08			
denovo97	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Dorea; s__formicigenerans	-0.114 (0.022)	2.46E-07			
denovo414	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Roseburia; s__	-0.091 (0.022)	3.67E-05			
denovo79	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__; s__	0.097 (0.022)	8.32E-06			
denovo123	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__; s__	0.103 (0.024)	2.83E-05			
denovo12	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__; s__	0.167 (0.025)	8.61E-11			
denovo1380	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__; s__	0.095 (0.022)	1.10E-05			

Table 1. List of OTUs associated with hippurate, the hippurate diet score and foods [1]

OTU ID [2]	Assigned taxonomy	Hippurate		Diet score		Foods [3]
		Beta (SE)	*P*	Beta (SE)	*P*	(*P*<0.05)
denovo438	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__; s__	0.100 (0.025)	5.01E-05			
denovo33	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__; s__	0.098 (0.024)	4.87E-05			
denovo469	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Faecalibacterium; s__prausnitzii	0.100 (0.023)	1.66E-05	0.034 (0.010)	9.24E-04	WG: 0.007(0.003)
denovo276	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Oscillospira; s__	0.110 (0.026)	2.53E-05			
denovo59	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Oscillospira; s__	0.170 (0.026)	2.17E-10			
denovo424	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Ruminococcus; s__	0.107 (0.026)	3.36E-05			

*= statistically significant: *P*<0.0017; WG: whole grain products; OTUs, operational taxonomic units

(4) Microbiome OTUs significantly associated with both hippurate and the hippurate diet score are shown. Associations were adjusted for covariates (age, Shannon Index, metabolite batch, BMI, sex and family relatedness) and multiple testing using Bonferroni correction. Hippurate diet score associations were also adjusted for hippurate.

(5) OTU ID assignment is specific to the TwinsUK cohort.

(6) All foods included in the hippurate diet score were fitted into a backwards stepwise linear regression using *P*<0.05 as the cut-off threshold with each collapsed taxa/OTU associated to both hippurate and the diet score. Results displayed are the betas with standard errors of foods at least nominally associated (*P*<0.05). Statistical significance was defined as *P*<0.0017 (Bonferronil: 0.05/[10 taxa x 3 foods]).

Table 2. List of collapsed taxa associated with hippurate, the hippurate diet score and foods [(1)]

					Hippurate		Diet score		Foods
Phylum	**Class**	**Order**	**Family**	**Genus**	**Beta (SE)**	***P***	**Beta (SE)**	***P***	***P*<0.05 [(2)]**
Actinobacteria	Actinobacteria	Actinomycetales			-0.083 (0.022)	1.31E-04	-0.035 (0.011)	1.67E-03	Fruit: -0.004(0.002) WG: -0.007(0.003)
Actinobacteria	Actinobacteria	Actinomycetales	Actinomycetaceae		-0.089 (0.021)	2.89E-05	-0.036 (0.011)	1.70E-03	Fruit: -0.004(0.002) WG: -0.007(0.003)
Actinobacteria	Actinobacteria	Actinomycetales	Actinomycetaceae	*Actinomyces*	-0.101 (0.021)	1.55E-06	-0.045 (0.011)	5.71E-05	Fruit: -0.005(0.002) WG: -0.008(0.003)
Firmicutes					-0.102 (0.024)	1.94E-05			
Firmicutes	Clostridia				-0.088 (0.024)	2.44E-04			
Firmicutes	Clostridia	Clostridiales			-0.088 (0.024)	2.40E-04			
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae		-0.125 (0.022)	1.72E-08			
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	*[Ruminococcus]*	-0.111 (0.022)	4.03E-07	-0.038 (0.011)	6.35E-04	Fruit: -0.005(0.002) WG: -0.008(0.003)
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	*Blautia*	-0.146 (0.021)	5.68E-12			
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	*Dorea*	-0.116 (0.021)	6.13E-08			
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	*Roseburia*	-0.089 (0.022)	4.18E-05			
Firmicutes	Erysipelotrichi				-0.111 (0.021)	1.64E-07			
Firmicutes	Erysipelotrichi	Erysipelotrichales			-0.111 (0.021)	1.64E-07			
Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae		-0.111 (0.021)	1.64E-07			
Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	*[Eubacterium]*	-0.083 (0.021)	9.30E-05	-0.040 (0.012)	6.12E-04	Fruit: -0.004(0.002) WG: -0.010(0.003)*
Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	*Ralstonia*	-0.084 (0.021)	7.85E-05			

*= statistically significant: *P*<0.0017; WG: whole grain products

(1) Microbiome collapsed taxa significantly associated with both hippurate and the hippurate diet score are shown. Associations were adjusted for covariates (age, Shannon Index, metabolite batch, BMI, sex and family relatedness) and multiple testing using Bonferroni correction. Hippurate diet score associations were also adjusted for hippurate.

(2) All foods included in the hippurate diet score were fitted into a backwards stepwise linear regression using *P*<0.05 as the cut-off threshold with each collapsed taxa associated to both hippurate and the diet score. Results displayed are the betas with standard errors of foods at least nominally associated (*P*<0.05). Statistical significance was defined as *P*<0.0017 (Bonferronil: 0.05/[10 taxa/OTUs x 3 foods])

Table 3. Associations between MetS status and criteria categories with Shannon diversity, the hippurate gradient, the hippurate diet score and associated OTUs and taxa

MetS phenotype	Variable	Beta(SE)	*P*	R^2
MetS status (0, no; 1, yes)	Shannon diversity	OR: 0.744 (0.064)	6.35E-04	0.0105
	Hippurate trajectory	OR: 0.795 (0.082)	0.026	0.0054
	Actinomycetaceae family	OR: 1.397 (0.145)	0.001	0.0110
	Actinomycetales order	OR: 1.373 (0.140)	0.002	0.0100
	Actinomyces genus	OR: 1.310 (0.134)	0.009	0.0072
	Eubacterium genus	OR: 1.215 (0.111)	0.034	0.0038
	Ruminococcus genus	OR: 1.308 (0.128)	0.006	0.0075
	Ruminococcus genus OTU denovo55	OR: 1.373 (0.145)	0.003	0.0099
	Clostridiales order OTU denovo299	OR: 0.703 (0.063)	9.47E-05	0.0122
	Faecalibacterium prausnitzii OTU denovo469	OR: 0.732 (0.065)	4.33E-04	0.0095
BMI	Shannon diversity	-0.770 (0.151)	4.63E-07	0.0288
	Hippurate trajectory	-0.700 (0.149)	3.25E-06	0.0240
	Hippurate diet	-0.332 (0.154)	0.032	0.0053
	Actinomycetaceae family	0.393 (0.152)	0.010	0.0075
	Actinomycetales order	0.420 (0.153)	0.006	0.0085
	Actinomyces genus	0.431 (0.155)	0.005	0.0089
	Ruminococcus genus	0.407 (0.149)	0.006	0.0081
	Ruminococcus genus OTU denovo55	0.430 (0.151)	0.005	0.0091
	Clostridiales order OTU denovo299	-0.317 (0.147)	0.031	0.0048
	Faecalibacterium prausnitzii OTU denovo469	-0.487 (0.143)	7.26E-04	0.0115
HDL-cholesterol	Shannon diversity	0.096 (0.027)	4.42E-04	0.0126
	Hippurate diet	0.081 (0.027)	0.003	0.0095
	Ruminococcus genus OTU denovo55	0.092 (0.027)	8.46E-04	0.0123
	Clostridiales order OTU denovo299	0.073 (0.027)	0.007	0.0076
	Faecalibacterium prausnitzii OTU denovo469	0.056 (0.026)	0.035	0.0044
Triglycerides	Shannon	-0.122 (0.027)	4.97E-06	0.0213
	Hippurate	-0.077 (0.026)	0.003	0.0087
	Diet	-0.094 (0.027)	5.72E-04	0.0130
	Actinomycetaceae family	0.055 (0.025)	0.029	0.0045
	Eubacterium genus	0.061 (0.026)	0.019	0.0056
	Ruminococcus genus	0.072 (0.026)	0.005	0.0077
	Ruminococcus genus OTU denovo55	0.109 (0.026)	3.56E-05	0.0178
	Clostridiales order OTU denovo299	-0.115 (0.024)	2.38E-06	0.0194
	Faecalibacterium prausnitzii OTU denovo469	-0.079 (0.025)	0.002	0.0092

MetS, metabolic syndrome; HDL, high density lipoprotein; OTU, operational taxonomic unit; OR, odds ratio
Notes: A linear regression was performed using Shannon diversity, the hippurate trajectory, the hippurate diet and hippurate and diet associated taxa/OTUs as predictors of MetS status (adjusting for age, sex, and family relatedness) and each component adjusting for age, BMI (except for BMI), sex, and family relatedness. Statistical significance was defined as $P<0.05$.

Table 4. Percent variance in the association between the hippurate trajectory and MetS, BMI and triglycerides accounted for through each associated variable

MetS phenotype	Variable [1]	Hippurate trajectory		% variance through variable
		r^2_x [2]	r^2_{xy} [3]	
MetS status (0, no; 1, yes)	Hippurate trajectory	0.0054		
	Shannon diversity		0.0021	61.1
	Actinomycetaceae family		0.0040	25.9
	Actinomycetales order		0.0042	22.2
	Actinomyces genus		0.0042	22.2
	Eubacterium genus		0.0052	3.7
	Ruminococcus genus		0.0046	14.8
	Ruminococcus genus OTU denovo55		0.0038	29.6
	Clostridiales order OTU denovo299		0.0026	51.9
	Faecalibacterium prausnitzii OTU denovo469		0.0031	42.6
BMI	Hippurate trajectory	0.0288		
	Hippurate diet		0.0203	29.5
	Shannon diversity		0.0122	57.6
	Actinomycetaceae family		0.0215	25.3
	Actinomycetales order		0.0217	24.7
	Actinomyces genus		0.0211	26.7
	Ruminococcus genus		0.0223	22.6
	Ruminococcus genus OTU denovo55		0.0207	28.1
	Clostridiales order OTU denovo299		0.0199	30.9
	Faecalibacterium prausnitzii OTU denovo469		0.0186	35.4
Triglycerides	Hippurate trajectory	0.0087		
	Hippurate diet		0.0055	36.8
	Shannon diversity		0.0032	63.2
	Actinomycetaceae family		0.0077	11.5
	Eubacterium genus		0.0084	3.4
	Ruminococcus genus		0.0079	9.2
	Ruminococcus genus OTU denovo55		0.0063	27.6
	Clostridiales order OUT denovo299		0.0045	48.3
	Faecalibacterium prausnitzii OTU denovo469		0.0060	31.0

MetS, metabolic syndrome; OTU, operational taxonomic unit

(5) All variables were significantly (P<0.05) associated with each listed MetS phenotype.

(6) The proportion of the variance in each MetS phenotype explained by the hippurate trajectory after taking into account all covariates (age, sex, family relatedness and BMI [for triglycerides]).

(7) The proportion of the variance in each MetS phenotype explained by the hippurate trajectory after taking into account all covariates as in (2) and adjusting for the applicable variable.

Table 5. Percent variance in the association between the hippurate diet score and BMI, HDL cholesterol and triglycerides accounted for through each associated variable

MetS phenotype	Variable [1]	Hippurate diet score		% variance through variable
		r^2_{x} [2]	r^2_{xy} [3]	
BMI	Hippurate diet	0.0053		
	Shannon diversity		0.0029	45.3
	Hippurate trajectory		0.0021	60.4
	Actinomycetaceae family		0.0042	20.8
	Actinomycetales order		0.0042	20.8
	Actinomyces genus		0.0039	26.4
	Ruminococcus genus		0.0045	15.1
	Ruminococcus genus OTU denovo55		0.0038	28.3
	Clostridiales order OTU denovo299		0.0038	28.3
	Faecalibacterium prausnitzii OTU denovo469		0.0034	35.8
HDL	Hippurate diet	0.0095		
	Shannon diversity		0.0074	22.1
	Ruminococcus genus OTU denovo55		0.0073	23.2
	Clostridiales order OTU denovo299		0.0070	26.3
	Faecalibacterium prausnitzii OTU denovo469		0.0079	16.8
Triglycerides	Hippurate diet	0.0130		
	Shannon diversity		0.0100	23.1
	Hippurate trajectory		0.0097	25.4
	Actinomycetaceae family		0.0118	9.2
	Eubacterium genus		0.0118	9.2
	Ruminococcus genus		0.0120	7.7
	Ruminococcus genus OTU denovo55		0.0100	23.1
	Clostridiales order OUT denovo299		0.0087	33.1
	Faecalibacterium prausnitzii OTU denovo469		0.0104	20.0

MetS, metabolic syndrome; HDL, high density lipoprotein; OTU, operational taxonomic unit

(1) All variables were significantly (P<0.05) associated with each listed MetS phenotype.

(2) The proportion of the variance in each MetS phenotype explained by the hippurate diet score after taking into account all covariates (age, sex, family relatedness and BMI [for triglycerides]).

(3) The proportion of the variance in each MetS phenotype explained by the hippurate diet score after taking into account all covariates as in (2) and adjusting for the applicable variable.

Appendix F. Chapter 7 Appendices

Table 1. Nutrient profile and linear trends of the visceral fat mass diet score according to score tertile

	Tertile 1	Tertile 2	Tertile 3	Trend	
Nutrient	Mean (SD)	Mean (SD)	Mean (SD)	Beta (SE)	*P*
Energy (kcal)	1892.1 (510.9)	1804.2 (522.9)	1870.4 (552.2)	-15.61 (14.56)	NS
Fat (g/d)	61.5 (10.4)	68.6 (9.5)	72.4 (10.3)	5.53 (0.28)	$1.62x10^{-75}$
Saturated FAs (g/d)	21.3 (5.4)	24.5 (5.2)	26.6 (5.1)	2.65 (0.14)	$1.90x10^{-67}$
MUFAs (g/d)	19.7 (3.7)	22.7 (3.4)	24.7 (3.8)	2.54 (0.1)	$2.33x10^{-113}$
PUFAs (g/d)	15.2 (4.4)	15.9 (4.4)	15.7 (4.5)	0.3 (0.12)	$1.21x10^{-2}$
trans-FAs (g/d)	1.3 (0.6)	1.6 (0.6)	1.8 (0.6)	0.25 (0.02)	$2.55x10^{-51}$
Cholesterol (mg/d)	202.5 (69.1)	234.8 (74.2)	251.5 (78.2)	24.98 (2.06)	$4.04x10^{-32}$
Protein (g/d)	79.5 (12.2)	81 (12.3)	79.6 (12.5)	0.15 (0.34)	NS
Carbohydrate (g/d)	248.1 (29.6)	229 (29.1)	217.2 (31.1)	-15.66 (0.84)	$1.13x10^{-69}$
Starch (g/d)	107.7 (26.3)	107.9 (25.1)	111.2 (26.1)	1.65 (0.71)	$2.05x10^{-2}$
Total sugars (g/d)	137.5 (28.1)	118.3 (24.4)	103.4 (24.1)	-17.22 (0.71)	$5.34x10^{-108}$
Glucose (g/d)	28 (9.2)	22.4 (7.8)	18 (6.5)	-5.05 (0.22)	$8.20x10^{-101}$
Fructose (g/d)	33.7 (11.4)	26 (9.2)	19.9 (7.5)	-6.99 (0.26)	$1.29x10^{-125}$
Sucrose (g/d)	48 (14.1)	44.2 (14.4)	42.7 (15.6)	-2.73 (0.41)	$6.34x10^{-11}$
Maltose (g/d)	3.3 (1.5)	3 (1.6)	3.3 (1.8)	-0.02 (0.05)	NS
Lactose (g/d)	18.7 (9.9)	18.3 (8.9)	17 (9.8)	-0.82 (0.26)	$1.82x10^{-3}$
NSP (g/d)	22.9 (5.5)	19.6 (4.5)	16.4 (4.2)	-3.22 (0.13)	$2.27x10^{-110}$
Alcohol (g/d)	8.4 (11.1)	8.8 (11.3)	10.9 (14.9)	1.24 (0.38)	$9.45x10^{-4}$
Water (g/d)	2754.6 (662)	2593.7 (598.9)	2478.5 (598.1)	-139.45 (16.71)	$1.80x10^{-16}$
Sodium (mg/d)	2308.6 (498.6)	2262.8 (471.7)	2162.2 (453.5)	-71.5 (12.87)	$3.33x10^{-8}$
Potassium (mg/d)	4080.4 (624.4)	3833 (569.2)	3550.3 (524.2)	-263.97 (15.56)	$1.48x10^{-58}$
Chloride (mg/d)	3705.3 (764.9)	3598.8 (728.6)	3411.9 (694.5)	-144.23 (19.7)	$4.30x10^{-13}$
Calcium (mg/d)	1095.7 (286.9)	1039.4 (260.8)	949.3 (264.1)	-72.17 (7.31)	$3.18x10^{-22}$
Magnesium (mg/d)	369.1 (52.3)	336.3 (49.2)	296.7 (42.4)	-36.01 (1.27)	$3.72x10^{-138}$
Phosphorous (mg/d)	1534.9 (218.8)	1474.9 (202.7)	1357 (211.8)	-87.14 (5.81)	$4.40x10^{-47}$
Iron (mg/d)	13.5 (3.1)	12.5 (2.4)	11.5 (2.3)	-0.99 (0.07)	$2.43x10^{-39}$
Copper (mg/d)	1.6 (0.4)	1.6 (0.5)	1.5 (0.5)	-0.07 (0.01)	$1.90x10^{-8}$
Zinc (mg/d)	10.1 (1.5)	10.1 (1.6)	9.8 (1.7)	-0.13 (0.04)	$2.60x10^{-3}$
Manganese (mg/d)	4.4 (1.1)	4 (1)	3.6 (0.9)	-0.43 (0.03)	$9.54x10^{-52}$
Iodine (ug/d)	212.9 (73)	207 (68.7)	196.5 (68.7)	-8.03 (1.89)	$2.25x10^{-5}$
Retinol (ug/d)	407.1 (403.2)	533.9 (586.6)	617.4 (772.6)	106.49 (16.78)	$3.05x10^{-10}$
Carotene (ug/d)	6060.1 (4193.3)	5399.6 (3002.5)	4729.8 (2742.7)	-664.87 (94.64)	$3.41x10^{-12}$
Vitamin D (ug/d)	2.5 (1.1)	2.5 (1)	2.4 (0.9)	-0.05 (0.03)	$5.36x10^{-2}$
Vitamin E (mg/d)	11.7 (3.4)	10.9 (2.9)	9.9 (2.8)	-0.92 (0.08)	$1.02x10^{-26}$
Thiamin (mg/d)	1.8 (0.4)	1.7 (0.3)	1.6 (0.3)	-0.12 (0.01)	$3.38x10^{-31}$
Riboflavin (mg/d)	2.3 (0.7)	2.2 (0.6)	2.1 (0.6)	-0.09 (0.02)	$2.03x10^{-7}$
Niacin (mg/d)	21.7 (5)	21.1 (4.5)	20.5 (4.5)	-0.59 (0.13)	$4.09x10^{-6}$
Tryptophan (mg/d)	16.6 (2.5)	16.9 (2.5)	16.6 (2.5)	0.02 (0.07)	NS
Vitamin B6 (mg/d)	2.6 (0.6)	2.5 (0.5)	2.3 (0.5)	-0.15 (0.01)	$2.63x10^{-23}$
Vitamin B12 (mg/d)	5.8 (2.2)	6.3 (2.5)	6.4 (2.9)	0.31 (0.07)	$6.65x10^{-6}$
Folate (ug/d)	416 (116)	378.8 (103.5)	341.8 (91.7)	-37.1 (2.79)	$5.95x10^{-38}$
Pantothenate (mg/d)	6.1 (1.6)	6 (2.8)	5.6 (1.3)	-0.23 (0.04)	$1.43x10^{-8}$
Biotin (ug/d)	46.9 (10.1)	45.5 (9.4)	42.3 (9.5)	-2.27 (0.26)	$1.50x10^{-17}$
Vitamin C (mg/d)	198.9 (81.5)	164.1 (66.9)	125.8 (50.8)	-36.44 (1.82)	$8.20x10^{-78}$

NS= not significant: P>0.05. FA: fatty acid, MUFAS: monounsaturated fatty acids, PUFAS: polyunsaturated fatty acids, NSP: non-starch polysaccharides

Tertile 1: score 0-6; tertile 2: score 7-9; tertile 3: scores 10-15. Linear trend determined by using the tertile of the VFM diet score as a predictor of the energy-adjusted nutrient intake.

Table 2. Associations between visceral fat mass (VFM) and the VFM diet score, Shannon Index and top microbiome and metabolite associations in the MZ discordant twin sample

		Low-VFM twins	High-VFM twins	Regression VFM		
Variable	N pairs	Mean(SE)	Mean(SE)	Beta(SE)	*P*	R^2
VFM diet score	80	-0.141(0.122)	-0.047(0.115)	0.281(0.091)	0.002	0.057
Alpha-hydroxyisovalerate	80	-0.239(0.120)	0.064(0.111)	0.141(0.094)	0.136	0.014
Bilirubin (Z,Z)	68	0.093(0.102)	-0.110(0.134)	-0.187(0.102)	0.068	0.023
Butyrylcarnitine	80	-0.168(0.126)	0.096(0.124)	0.199(0.087)	0.023	0.032
Hippurate	80	0.062(0.112)	-0.162(0.114)	-0.297(0.095)	0.002	0.058
Shannon Index	27	-0.019(0.182)	-0.268(0.224)	-0.287(0.176)	0.110	0.048
Bifidobacterium OTU (4426298)	27	0.140(0.718)	0.175(1.098)	-0.064(0.197)	0.749	0.002
Eubacterium dolichum	27	-0.108(0.155)	-0.013(0.171)	0.131(0.227)	0.566	0.006

Notes: A linear regression was conducted using the VFM diet score, Shannon Index and top microbiome and metabolite associations to predict VFM in the MZ discordant (1 SD apart in VFM) twin sample. Variables standardized to have mean=0, SD=1.

Appendix G. Publications throughout my PhD

Pallister, T., Spector, T.D. **Food: a new form of personalised (gut microbiome) medicine for chronic diseases?** Sep 2016 In: *Journal of the Royal Society of Medicine*. 109, 9, p. 331-336.

*I wrote this review in a joint effort with Professor Tim Spector.

Pallister, T., Haller, T., Thorand, B., Altmaier, E., Cassidy, A., Martin, T., Jennings, A., Mohney, R.P., Gieger, C., MacGregor, A., Kastenmüller, G., Metspalu, A., Spector, T.D., Menni, C. **Metabolites of milk intake: A metabolomic approach in UK twins with findings replicated in two European cohorts.** 28 Jul 2016 In: *European Journal of Nutrition.* [Epub ahead of print]

*Work for this publication appears in Chapter 4.

Pallister, T., Jennings, A., Mohney, R. P., Yarand, D., Mangino, M., Cassidy, A., MacGregor, A., Spector, T. D. & Menni, C. **Characterizing blood metabolomics profiles associated with self-reported food intakes in female twins.** 29 Jun 2016 In: *PL o S One.* 11, 6, e0158568

*Work for this publication appears in Chapter 4.

Pallister, T., Sharafi, M., Lachance, G., Pirastu, N., Mohney, R. P., MacGregor, A., Feskens, E. J. M., Duffy, V., Spector, T. D. & Menni, C. **Food Preference Patterns in a UK Twin Cohort**. Dec 2015 In: *Twin Research and Human Genetics*. 18, 6, p. 793-805

*I created an online food preference questionnaire based on Dr Valerie Duffy's questionnaire. I used this questionnaire for validating the food metabolite scores in Chapter 5.

Pallister, T., Spector, T. D. & Menni, C. **Twin studies advance the understanding of gene-environment interplay in human nutrigenomics.** Dec 2014 In: *Nutrition Research Reviews*. 27, 2, p. 242-251

*I wrote this literature review which appears in Chapter 1.